

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 14:49:40 : Search time 42.14 Seconds
(without alignments)
3642.714 Million cell updates/sec

Title: US-09-729-653-2
Perfect score: 7257
Sequence: 1 HSLIGRCSRGILGDGNNAVAC.....AAOLKKISDTQYLRTHEEL 1382

Scoring table: BLASTSUM62
Gapop 10.0, Gapext 0.5
Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: A.Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2780	38.3	1366	22	ABR65570 Drosophila melanog
2	2623	36.1	515	22	AAH80287 Human protein SEO
3	2544	35.1	502	22	AAH79303 Human protein SEO
4	1229	16.9	244	22	AAU14657 Novel bone marrow
5	1163	16.0	237	22	AAU14657 Novel bone marrow
6	706	9.7	138	22	AAU14751 Novel bone marrow
7	298	4.1	109	22	AAU13123 Novel human secret
8	293	4.0	1749	22	AAH79130 Human protein SEO
9	289	4.0	1205	22	ABR58628 Drosophila melanog
10	274.5	3.8	72	22	AAU33013 Novel human secret
11	263	3.6	1241	22	AAH79062 Human protein SEO

12	260.5	3.6	1252	21	AAH43197 Human ORFX ORF2961
13	259.5	3.2	1690	21	AAH42248 Human ORFX ORF2012
14	235.5	3.6	993	22	AAE01339 Human gene 24 enco
15	235.5	3.2	993	22	AAH93308 Human protein sequ
16	235	3.2	1443	22	ABH68472 Drosophila melanog
17	216.5	3.0	864	22	AAH94031 Human protein sequ
18	213	2.9	942	22	AAH80046 Human protein sequ
19	206.5	2.8	873	22	AAH95245 Human protein sequ
20	206	2.8	821	22	AAH93491 Human protein sequ
21	174.5	2.4	968	22	ABH63037 Drosophila melanog
22	170	2.3	887	22	ABH68631 Drosophila melanog
23	167	2.3	1332	20	AAV01084 Human IKAP protein
24	167	2.3	1332	22	AAH66345 Human IKAP protein
25	166	2.3	1332	22	AAH93160 Human ORFX ORF3013
26	164	2.3	496	21	AAH43249 Human polyprotein
27	161.5	2.2	1246	22	AAH39120 Human polyprotein
28	157	2.2	845	21	AAH43631 Arabidopsis thalia
29	157	2.2	923	21	AAH43630 Arabidopsis thalia
30	156.5	2.2	2998	18	AAH19691 ATM mutant 3403del
31	155	2.1	1270	22	AAH99141 Pyrococcus abyssi
32	155	2.1	1455	22	AAH99140 Putative P. abyssi
33	155	2.1	1458	22	AAH96825 S. pombe Rad3 poly
34	154.5	2.1	2386	18	AAH13153 Novel human diagno
35	154	2.1	769	22	AAH13410 Human polyprotein
36	150.5	2.1	1294	22	AAH40906 Amino acid sequenc
37	149.5	2.1	475	21	AAH53985 Apaf-1XL/Delta2-10
38	147.5	2.0	1149	22	AAH97640 Apaf-1MD protein s
39	147.5	2.0	1237	22	AAH97638 Apaf-1XL protein s
40	147.5	2.0	1248	22	AAH97639 Apaf-1XL-M368L pro
41	147.5	2.0	1248	22	AAH97641 Apaf-1XL-D10A prot
42	147.5	2.0	1248	22	AAH97642 Apaf-1XL-D27A prot
43	147.5	2.0	1248	22	AAH97643 Apaf-1XL-R390 prot
44	147.5	2.0	1248	22	AAH97644 Apaf-1XL-R630 prot
45	147.5	2.0	1248	22	AAH97644

ALIGNMENTS

RESULT 1
ID ABB65570 standard; Protein; 1366 AA.
XX ABB65570:
AC
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 23502.
DE
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmacological.
KM
XX Drosophila melanogaster.
OS
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
PF
XX 23-MAR-2001; 2001MO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
DR WPI: 2001-656860/75.
DR N-PSDB; ABL09673.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

DAVIS M.
09/1724653
Seq. ID 2

```

XX PS Disclosure: SFO ID NO 23502; 21bp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins.
CC sequences (AB101840-AB16175) and the encoded proteins.
CC (AB101840-AB16175).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1366 AA;

```

Query Match 38.3%; Score 2780; DB 22; Length 1366;
 Best Local Similarity 40.8%; Pred. No. 7, 6e-241;
 Matches 553; Conservative 286; Mismatches 468; Indels 50; Gaps 14;

```

OY 57 IGFAMKTSIGNVLAATGADYIVKIFEDRHGQKSEINLPNCVAMDMKDGVLAVIAEKS 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 19 vyfiwqka--llattgtdgsvalynrgqjlvqrlilsglscgfawdgedglilgtssg 75
OY 117 SCITVMDAFTNKTSOLDNGMRDOMSFLMSKVSFLAVGTVGKLNKITYNHQT-SRKIPVL 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 76 pulctwldyngskisveitgrlplctcllwsqqlavtgrfgnqalalyhssgskrpt 135
OY 176 GKHTRTTCGCNNAENLXALGEGEDKMITVSNOEGPTIRTOYRSEPPXNQOFLMKRDDPT 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 136 ghsrkritcgasagqllalgedsksfisnedgdtvrvqrlrdaptdmyfaemanderi 195
OY 236 SAAESMISVVLKKTFLFNLNEPDNPADLEFQDGFNIVCYNMYGGRIMTGFSCGHFV 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 196 -agdaismilgkrtfllylpepenptelgfgsrysglmqhkfwfdgyillgfsnghv 254
OY 296 VISTHTGELGGEIFQARNKNDLTSIAVSQTLNKVATGDCNCKIKIDDLVLDKMYIYIL 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 255 aistripkdvqgelwqykmdsklctglaycpcldivascgddsklnstlnbetercliv 314
OY 356 DEENKGLGTLSTWDGQLALSTORSLHVFLLKLPILLDACSTRIVLTSLEVTV--- 412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 315 -pdhagvqmldspdgqllavetlnhtylyvtklphlavesaprlivllsslaevslyy 373
OY 413 ANPVGELPITYSVDEVPFVAVGLYHLAVGKNNRAMFVLEGN--AVKKIKMEYLGCT 469
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 374 apdktsklsiprlpgegrflmavpynfatglekhwfydlgkslgeepriiserdiprs 433
OY 470 VASICLHSDYAALFEGKVOHLIESELDAQEBRET---RLFPVDDKCRILCHALITSD 526
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 434 vesmlinaaycaalcoppqlllqalaadnpckklqavfrtalpmpsdavltcltlasge 493
OY 527 FLIVGTIDGVGVVYFIEDMVFVNDYRHRPVSKKIFPPDNGTRLAVFIDEKSDGFVYCPVND 586
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 494 llifatlghlvtfyleakwdsctlyrhsmgtrqlfmdlsgtvlfidhsgsyvflpvve 553
OY 587 ATVEIPDSPTTKGVLMENKMPDKGVFLAYDDDKVYTYFHKDTIQGAKVILLAGSTKYPE 646
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 554 eaalliprdipkqclglwd--lcpnlflisyladarivnthfvfshvsgtchtlmvgesklmp 611
OY 647 AHKPLLNGELTCQTOSGKVNNTYLSHGFLSLNKDKSPDLRLRMALNMLKRFSDMW 706
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 612 gqflllccggemalldggqyatqstlch--vvnpnsnqaanlqnl---lllryndey 665
OY 707 EMCRLINDEAAMNELARACLHMEVEFAIRVRRIGNVGIVSLDQIKIEDYVNLACHL 766
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 666 klckimgssawrefgegaiedlepdialtrqqlgdaalvalaselyvelldmlngcc 725
OY 767 AMFTNDVYLAQDLYLASSCPALAEKMRDLQHWSALDLAKHLADQQLPFISKEYAIOLE 826
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 726 clllaqydaqekemlkgyrraaalclrdlllqwdqalllanhndqgevytaryaqgle 785

```

```

OY 827 FAGDYVNAIHEKGTTCG-----NKEHDEACTAGIAQMSIRMGDIR 868
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 786 fngnycdalyhekgykedlinsketetdaImdspegeehvrlcmjgartsigedfr 845
OY 869 RGNVALKTHPSRYLAKKNDGCAILENMKQFSPAQLYKKGILYOKAASVYIRSKMMAVGLD 928
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 846 rlgayaveledqllldccellatvghlteaaglyer99fydeagcnytalmmvannanl 905
OY 929 LPHVSSPKIHLQYAKAKREADGRYKREAVVAYENAKQMSYIRYLDLNNPERAVNIVRET 988
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 906 lpkvsktllhaayakakendgnyheaelrsyrlagldidacvryllldhlcphaaesvlves 965
OY 989 QSLDGAKMVARFFLOLDGYSALQIFLVMSKCNNEATTLAQHKKMITYADITGS-EDTIN 1047
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 966 rsmdsakllaktyklygdvgeqalqlflvsgcveeaalagrkmlrthgellieryenaks 1025
OY 1048 EDYOSIALYEESEKRYLQKGFLLCGOYSRALKHFLLKCP--SSEDNVAVIEMAVTGVGA 1105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1026 sdylalahyfegekyllyagkyflareftkalrlllkasafmneeqvsistalcvats 1085
OY 1106 KDELTLNOLIDLLENDGMPKDAKYLFLYALOKYREAAQTALLIAREEOSAGNYRNA 1165
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1086. nneqatqliefllgevdgvpdkpyrlfrylmarrkhykdaaktavliandeqlagnykea 1145
OY 1166 HDVPSMVAELSQKIKIPSEVATNMLIHSYILVKIHVKNQDHMKGAKMLIRVANNISK 1225
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1146 rdllysmgygelrrnlsavtaemrhgfillhyrlylvihvklgnhllaakllqvaaacisq 1205
OY 1226 PFSHVPIILTSVYIEBCHRGKNSAFSFAMLMREPRFSKIDAKYKKIEGVWRPDP--- 1282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1206 fhelhlplltsevleehragllksaflyastlmrpdryngldpryakkleslyvrkapkrl 1265
OY 1283 ---ISEIEEATTPCEPKFLPECELLCPGCKNSIPYCIATGRHMLKDMWYCPHCDPFA 1339
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1266 kglrleidetmcepcldenslanmewtvcyscktlpiclataghlkqmtscpqcdfic 1325
OY 1340 LYSELKIMLNTSESTPCMCSERLMAOLKKISDCTOYL 1376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1326 fraemenlisengecpmcgenavapeqlldavedirpyl 1362

```

RESULT 2
 AAM80287
 ID AAM80287 standard: Protein; 515 AA.
 XX
 AC AAM80287;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human protein SEQ ID NO 3933.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 OS Homo sapiens.
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.
PA
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR N-PSDB; AAK53420.
XX
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
PS Claim 20; Page 471; 6221pp: English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK60302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM0020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 515 AA:
SQ
Query Match 36.1%; Score 2623; DB 22; Length 515;
Best Local Similarity 99.6%; Pred. No. 1.9e-227;
Matches 508; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
OY 797 QHMSALQAKHLAPDIPFISKEYAIQLEFAGDYVALAHYKGTGDNKEHDEACIAG 856
Db :|||||
6 ehwsaqlqakhlapodipfiskayaqlfagdyvalahykgldgdnkeheactag 65
OY 857 VAQMSIRMGDIRGCVNOLKHPSVRLKRDGCAILENNKQSEAAQLYEKGLYDKAASVY 916
Db :|||||
66 vaqmsirmgdirgcvnolkhpssrvlkrdcgallennmkqfseaaqlyekglyvkaasyv 125
OY 917 IRSNNAKVGDLPHVSSPKRHLOYAKAKKEDGKKEAVVAYENAKQOSYIRIYLDHLN 976
Db :|||||
126 irsnnakvgdldphvssprkhlolyakakkeadgrykaavvayenakqgsvirlyldhln 185
OY 977 NPEKAVNIVRETOSLDCAKMYARFELQDYGSAIOFLVMSKCNNEAFTLAQOHKKMEIY 1036
Db :|||||
186 npekaavivretosldcakmyarfflqldygysaioflvmskcnneafllaqohkmeiy 245
OY 1037 ADIGSEDTNEDYQSIALYFEGEKRYLQACKFELCGQYSRALKHLKCPSSDVAIE 1096
Db :|||||
246 adigsedtneqysialyfegekrylqagkfllcgysralkhlkcpssedvaie 305
OY 1097 MAITVQAKKELLTNQLIDHLGENDGMPKDAYLRLYVALKQYRAOTAIIRAREE 1156
Db :|||||
306 maitevqakelltnqlidhlgendgmpkdaylrllyvalkqyreaotaiiraree 365
OY 1157 OSAGNRYRAHDVLFMSVAELSKQIKIPSEMATNMLIHSTYILVKIHKNNDHMKGAMTL 1216
Db :|||||
366 osagnryrahdvlfmsvaelskqikipsematnmlihstylvkikhndhmkgaml 425
OY 1217 IRVANNISKPPSHIVPILTSTVIECHRAGLKNSAFSFAAMLMPREYSKIDAKYKKKIEG 1276
Db :|||||
426 irvanniskppshivpiltstviechraglknasfsfaamlmpreysskidakykkkieg 485
OY 1277 MYRRPDISEIEATTPCFCKFLPECCFL 1306
Db :|||||
486 myrrpdiseieattpcfckflpeseell 515

RESULT 3
AAM79303
ID AAM79303 standard; Protein: 502 AA.
XX
XX
XX AAM79303:
XX
XX
XX 06-NOV-2001 (first entry)
DE
XX Human protein SEQ ID NO 1965.
XX
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
OS
XX
XX W0200157190-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX
XX 05-FEB-2001; 2001WO-0504098.
PF
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX
XX (HYSE-) HYSEQ INC.
XX
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR N-PSDB; AAK52436.
XX
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
PS Claim 20; Page 4376-4377; 6221pp: English.
XX
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK60302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM0020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 502 AA:
SQ
Query Match 35.1%; Score 2544; DB 22; Length 502;
Best Local Similarity 99.0%; Pred. No. 2.5e-220;
Matches 493; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 708 MCRITNDAANNEALARACILHMEVFATRYRRIGNVQIWSLEIKIETEDYNLLAGILA 767
Db :|||||
1 mcritndaannealaracilhmevfatryrrignvqiwslleikietedynllagila 60

Oy	768	MFNDNYNLAADOLYASSCPIPLAAEMRBDLOHMSAOLAKHILAPDOIPLRSKEYAOLEE	827
Db	61	mfndnylnaagqlYlasscpplaalemtrdqlhwdsalqlakhlapdqjPltfskeyaqlEf	120
Oy	828	AGDYVNALAHKEKITTGDNKEHDEACLAGYAOMSIRMGDIRRGVNOLAKHPSKVLKRDGC	887
Db	121	agdyvnalahyekjItgdnkendeaacIagyaqmsIrmgdltrvgqaklhpstvlkrdcg	180
Oy	888	AILENKKOFSSAOLYKGLYYDKAASVYTRSKNMAKVGDLPHVSSPKIHLQYAKAKTA	947
Db	181	ailemkkofseaaqIyekjIydkaaasvytrsknmakvqglLphvsspkIhlyakakaa	240
Oy	948	DGRKEAVVAENKKOMOSYRIRYLDHLNPEKAVNVIRETOSLDQAKVAREPFLDGLGY	1007
Db	241	dgrkeavvayenakqmgsvytrIyldhlmpkekavnviretgsldqakmvarfIldgdy	300
Oy	1008	GSAIOFLVMKSCNNDEAPFLAQQHMKMEIVADIIGSEDTTNEDYQSIALYFEGEKRYLQAG	1067
Db	301	gsaiqlfvmkskcneeaflIaqhmkmeIyadiIgsedttneqysIalyfegekrylqag	360
Oy	1068	KFFLLCGYSALKHPLKCPSEEDNVAILMAIEVGAQKELLTNLOLHILGENDGMXK	1122
Db	361	kffllcgysalkhplkcpseednvailmaIevgaqakelltnqlldhIlgendsmxk	420
Oy	1128	DAKYLFRLYMALKOYREAAQTAAIIIAEEDOSAGNYRNAHDVLFSPYAEILKSOKIKIPSEM	1187
Db	421	dakylfrlymalkyreaaqtaaiIiaeeqsaagnynahdvlfsmyaeksqkIkpsem	480
Oy	1188	ATNLMILHISYILVAKIHVK 1205	
Db	481	atnmlilhysyllvfmIk 498	

XX AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequences may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins
CC and in assays to identify modulators of their expression and activity.
CC The anti-bone marrow protein antibodies and antagonists may also be used
CC to down regulate expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins
CC may be used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).
XX

Query Match	16.9%	Score 1229;	DB 22;	Length 244;
Best Local Similarity	96.3%	Pred. No. 4e-102;		
Matches 234; Conservative	0;	Mismatches 3;	Indels 6;	Gaps 1;

QY	100	MDMCKDEGVLAIVIAEKSSCIYLMADANTNKTSQLDNGRQDMSPFLMWSKVSFLAVGVK	159
Db	1	mdwdkdgdlavlaieekssciylwdanlnktsqldngmrqdmfllwskvysflavglvkg	60
QY	160	NLXIYNHOTSRRIPVLGKHTTRICGCGMAENLALGEGDKMITVSNQEGDTRIRQVR	219
Db	61	nlllynhbtsrktkipylyghktrfcgcmnaenllaiggedkmltvsnggedtlrqqvrs	120
QY	220	EPXNMQFLMKMDRTSAESMISVVLGKKTLPFLINPEPDNPADLEFQDFGNI-----	274
Db	121	epsmnqgffllmkmddrttsaesmlsvvlgkktlfflinpepdnadlefgqdfgnlvcym	180
QY	275	-VCYWMWEDGRIMIGFCGGRFVVISHTHGEIGDEIFDARHNKDNLSIAVSQTLNKVATC	333
Db	181	yvcymwygdgrlmigfscghfvstlhtlgelgelfgarnhkdnltaivsqtlmkvate	240
QY	334	GDN 336	
Db	241	gdn 243	
RESULT 5			
ID	AAG74711		
XX	AAG74711 standard; Protein; 237 AA.		
AC	AAG74711;		
DT	03-SEP-2001 (first entry)		
DE	Human colon cancer antigen protein SEQ ID NO:5475.		
KM	Human: colon cancer; colon cancer antigen; diagnosis; detection;		
XX	colorectal carcinoma; chromosome 1.		
OS	Homo sapiens.		
PN	WO200122920-A2.		
PD	05-APR-2001.		

PF 28-SEP-2000; 2000MO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX WPI; 2001-235357/24.
DR N-PSDB; AAH34116.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 7076-7077; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 237 AA:

Query Match 16.0%; Score 1163; DB 22; Length 237;
Best Local Similarity 99.1%; Pred. No. 3,4e-96;
Matches 228; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1050 YGATATFEEGKRYLQAGKFFELLCGOVSRAIKHFLKCPSSSDNVAITMAIETVGOADEL 1109
DB 1 YGATATFEEGKRYLQAGKFFELLCGOVSRAIKHFLKCPSSSDNVAITMAIETVGOADEL 60
QY 1110 LTNQLDHLGENDGMPKDAKYLFRLYMALKOYREAAQTAAIITAREBOSAGNKNADVL 1169
DB 61 ltnqldhlhgendgmpkdkylfrlymalkyreaaqtaalllareeqxagnynadvl 120
QY 1170 FSNYAEKSKRKIRIPSEMATNMLNHSYIIVKTHVKNQGMHMKARMLIRVANNISKPPSH 1229
DB 121 fsmnyeakskqkikipsematnmlnhsyilvkthvknqgmhmkarmlirvaanniskpsh 180
QY 1230 IVPILFTVTECHRAGLKNASFAAMLMRPEVRSKIDAKYKKKIEGMVR 1279
DB 181 lvpilftvtechraglknasafaaamlmrpeyrskidakykkkkiegmvr 230

RESULT 6
AAU14751
ID AAU14751 standard; protein; 138 AA.
XX
AC AAU14751;
XX
DT 24-OCT-2001 (first entry)
XX
DE Novel bone marrow polypeptide #150.
XX
KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;

KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.
XX
XX Homo sapiens.
XX
XX WO200157187-A2.
XX
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001MO-US03782.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 30-NOV-2000; 2000US-0250683.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Ford JE, Boyle BJ, Tang YF, Liu C, Asundi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
XX
XX WPI; 2001-488875/53.
DR N-PSDB; AAS23056.
XX
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT and gene therapy -
XX
PS Claim 10; Page 135; 392pp; English.
XX
XX AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequences may also be used as
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins
CC and in assays to identify modulators of their expression and activity.
CC The anti-bone marrow protein antibodies and antagonists may also be used
CC to down regulate expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins
CC may be used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ Sequence 138 AA:

Query Match 9.7%; Score 706; DB 22; Length 138;
Best Local Similarity 98.5%; Pred. No. 2,4e-55;
Matches 132; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 169 SRKIPVLGKHTKRTTCGMAENLXALGGEKMTTVSNOEGDTRQTVSEPNMOFFL 228
DB 5 srkipvlgkhtkrttcgmaenlxlalggdekmtvtvsnoegdtrqtvsepsnmffl 64
QY 229 MKMDRTSAEEMISVVLGKRTLEFLNLNEPDNPADLEFOODFGNIVCYNMGDGRIMIG 288
DB 65 mkmdrtsaeemisvvlgkrtleflnlnepdnpadlefqdfgnivcynmgdgrimig 124
QY 289 FSCGHFVIVSTHNG 302
DB 125 fscghfvivsthtg 138

RESULT 7
 ID AU033123 standard; Protein: 109 AA.
 AC AU033123;
 DT 18-DEC-2001 (first entry)
 DE Novel human secreted protein #3614.
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 OS Homo sapiens.
 PN MO200179449-A2.
 XX
 XX
 PD 25-OCT-2001.
 PF 16-APR-2001; 2001MO-US08656.
 PR 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 PA (HYSE-) HYSEQ INC.
 XX
 XX
 PI Tang YT, Liu C, Drmanac RT;
 DR WPI: 2001-611725/70.
 XX
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 XX vaccination, testing and therapy -
 XX
 XX
 PS Claim 20; Page 710; 765pp; English.
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemia. AA029510-AA03304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX
 XX Sequence 109 AA;
 XX

```

OY 157 VKGNLXIV--NHQSRKIPVIGK--HTKRITCGMNAENLXALG---GEDKMITVSNQEG 209
D 86 tdlhlyvkkigedwgdqkvcnkkfifqtsavtcjwpaeyilivfqlaegkvrlantknks 145
OY 210 DTRQTVRSEPNYKMOFLKMKDDRTSAESMISVVLGKKTLEFLNLNEPDNADLEFQO 269
D 146 stlygt-----eayvslctncsgkqlsgnadtivrf-----ddegsgsg 191
OY 270 DFGNIYC---YMWGDGRIMIGFSCGHFVYISTHNGELGCEIFQANHNHNDLSTAVSQ 325
D 192 klvnhpcrpypalawatslvaag--corklvaygkeghmlqtdydsrpbqrefltsvss 249
OY 326 TLAKVATCCG--NCIKIODLVDLKMVYILNDEENK-----GLGCTMTDQGLLAL 376
D 250 psgsgvvlgsydlrlyvnmwprtsiv-----eakrkeltlnlytalaawkdgsrlcv 303
OY 377 STGRSLHVEFLTKPLIGDASTRIAY-----LTSL--LEVYANPYEGELPITVS--V 426
D 304 gtlcgvgvf-----dcclrtslvknkfeltyvgpsqvivknlsqgrtvlkshygy 355
OY 427 DVERNFYAVGLYLAVGMNNRAMPYVLGENAVKKLDMELGVA-----STCL-- 475
D 356 eve-evklilgkerylvahlseet--lllgdlntrnlselawgsgnekyffenenvcmlf 412
OY 476 -----HSDYAALFEKGKQVLHLESEILDQOE---ERETRLFPVADDCRILCH 521
D 413 nageltlveymndltgsvtelfmphiivrlnercqrgteknklaylidlktiaivd 472
OY 522 ALTSDFLIGTDTGVVQFYIEDMOFVNDYRHPYSVKRIPDPNGRLVEIDKSDGFVY 581
D 473 -----llggyrlgtvshesgrvdlne-----tghkllfrdx----- 506
OY 582 CPVNDATYERDPSPT--IKGVLMMNPRMDKGFIAID-----DDKYTYTVFH 627
D 507 --lrlhlydiescsktilnlfcsymqvwpsdvlvagrnslcwuyiaepetvmtflir 564
OY 628 KDTI-----OGAK---VILAGSTKVPFAHKPLLLYNGELTQOTOSGVNNIYLSHGLSN 680
D 565 gdlvlglergsgklevmmegvltvayt-----ldegliefgtaldqgn--ylyrataflet 617
OY 681 LKDXGPDELRLPMLNLMLKRFSDAWEMCRILNDEAAMNELARACILHMEVEFAIRYVR 740
D 618 l-----emtrp-----eteamwktlskialearqlhaercfsa 650
OY 741 IGVNGIVMSLEQIKGT-----EDYNLAGHLAMFTNDYNIADQLIASSCPPIA 788
D 651 lgvvakarflhetneladqvsreygsgtdfyvtratlamlcknyklaemlfleqnaeve 710
OY 789 ALEMRRLQIHMDSALQALAKHILAPDQIFPISKEVAYIOL-----EFAGDYNAL 835
D 711 amgmvgelhwdeclavaeekyhpaleklttrsyqwlmdcqgeeragelgesgdglaai 770
OY 836 AHYKGTGDNKEHDEACLAGVAOMSTRMGDIRRGVNLKHSRVLKRDCGAILNMKKO 895
D 771 slvlt-----aglpakaarl-----vltre--ellantel 798
OY 896 FSEAAOLYKGLYDYDKAASYIRSKNMAKVGDLPHVSSKRIHQ-----YAKAE-A 947
D 799 vehltaailgelajera-----gdlfeklnpqkalecyrknaifakavela 845
OY 948 DGRYKEAVVAYENA-----KOMOSVIRLYDPLHNLNPKRANVIVETOSLIDGAKK--V 997
D 846 rlafveevvnlleaewghlvqgqldaaalnhye-----arclsiaaleaajargawk 898
OY 998 AREFLQGDYSAIOFLVMSKCNNEAFTLAOH---NKMETIYADITGSEDTTNEQOST 1053
D 899 alylaldtdntasky-----yplvaghyaslsgeyltlkgdrtkd--al 945
OY 1054 ALYFEGEKRLQAGKFFLLC-----GQYSRALKNHLKCCSSSEDNV 1093
D 946 dmytga-grwegahkklamkcmrpedvsvlyltqgemekgqkyreaerlyvtv--gepdl 1002

```

```

OY 1094 ALEM-----ALETVOAKDELTLNQLDHL-LIS---ENDGMPKDAKLF----- 1133
D 1003 aicmykhhklyddmirlyvghkhpollsd-----tlhlgkelaegrldgeayhylaegwk 1059
OY 1134 ---RLYWMALQYREAAQTAAITIAREEQAGNVRNADVLSMYAELKSQRIKIPSEMATN 1190
D 1060 atvmmyrasglweeay-----vartgga-----nahkhvaylwak----- 1096
OY 1191 LMLHSITLLVKIKHKNDDHKKGARMILIRVANNTSKFSPSHVPLITSTVIECHRAGLKNSA 1250
D 1097 -----slggeaavrllnklyg-----lleaavd-haa--dnscs 1125
OY 1251 FSEFAMLRPEYRSK---IDAKYKKRTEGWRPRDISIEEA 1289
D 1126 fefafelstlalkhktpevhkyamfie-----degkfeea 1161

RESULT 9
ABB58828
ID ABB58828 standard; Protein; 1205 AA.
XX
AC ABB58828:
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 3276.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN W0200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PDB; ABL02931.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 3276; 21pp + Sequence listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB5737-ABB72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX
SQ Sequence 1205 AA:

```

Query Match 4.08; Score 289; DB 22; Length 1205;
 Best Local Similarity 17.9%; Pred. No. 5.8e-16;
 Matches 261; Conservative 219; Mismatches 530; Indels 452; Gaps 62;

```

0Y 60 AMOKISGVYLLAVTGADYVVKFE-----DRHGOKRSEIINLPNGNVANDMKDQDVLAVIAE 114
Db 21 aaukeeg-y-lavagtdgdlkiklIdgatpugskvgglaavsn-lsmnqldgn-----k 72
0Y 115 KSSCIYILN-DANTNKTSTOLDNGM-----RDOMSEFLMSKVS--- 150
Db 73 esvrvvtvnddgqkrltsddcdgvlmwvmllykgsweemtdtrktsvsaamswtsgsrlic 132
0Y 151 -----FLAVGTVKGNIIXYHNSRTSRKIPVLGKHTK--RITGCGMAEN--LXALGED 199
Db 133 lvyedgalivsgvdgn-----lffgkelghlthlyqswspdnrlilfalaage 180
0Y 200 -----KNITV--SNEGDTI-----RQVQVSEFXNNOFLLMKMD 232
Db 181 chlYdngnlfamkllhigcvtlnlsgsgsgrtrtlaslcfwgrvvgstkrp----- 229
0Y 233 DRTSAEBSMISVVLGKTLFELFLNLEBPNADLEFOODFENIYCWYMGDGRIMIGFSG 292
Db 230 -----vliCengrvqgmrenddapa--lfdgmnnva-kwhdgcvla--lCg 276
0Y 293 HFV--VISTHNGELGQEIFQARNHK-----DNLTSIASOTLKNVATGCDNCIKIQ 341
Db 277 ttldavspLsgdrtnqvcfyfsprklytrllkvpgrdtltslwegkrltIamavdsffiy-- 334
0Y 342 DLVYDKMDYVILNDEENKGGTISWDDOQLALs--TQGSILHVTLYLP----- 391
Db 335 -faultrpdyIwcyfeklvflnsqsaIrespmviltwlvtsnqslfikevptlclasses 393
0Y 392 ---lIGDAC-STRIAYLTL-SLEETVANPVE-----GELPRTVS--VDVEPNFVA 434
Db 394 ehcvIagvcsvsnlkelaIstle-nrsnpddrvyqlllcnslgltcvskdyidrlpctfg 452
0Y 435 VGLNHLAVGMNRR--AMPYVLGENA-----VKKLDMEYLGTVASICLHSDYAALFEGK 487
Db 453 insyvvaIasgealllwyhtrpkascllhvkarkekrf-----hlldtprgvema 503
0Y 488 VOLNHESEIILDQEEERTLRFRAVDKCNILCHALTSPFLIYGTDPGVGVFYIEDNQF 547
Db 504 kdlmIsssgsqghstqr-----glscp--lcalalseklllvaresgaIneysIaanval 555
0Y 548 VNDYHNPVSVKRIEPPDPNGTFLYFIDKSDGFVUCVPNDVATYEIPFSPFIRIGVLME-NM 606
Db 556 nrthlmakvykkmIcnstaaIdlmg-vmtlIdlddnreqlnfsrvektvdawsw 614
0Y 607 PMDKGVFAIYDDDKVYVYVEHKDTIGAKVYILAGSTKVRPFAHKLPLLNGELTCOTGSGK 666
Db 615 atdnprllal-mektrmylfrgnpde-epvscsg-----yIctfedle 655
0Y 667 VNNIYLS---HGFSLNKKODGPELRLMLAH--NLMKRFSPAMCMKILNDEEA----- 717
Db 656 ltsvlllddlIsvvgelgn-----fshlqIrvksIrtlddlIehvylcdakqfi 703
0Y 718 -----WNLERACLIHMEVEFAIRVURRINCVGIVMSLEQIKIEDYNNLAGHLAF 769
Db 704 ednprhrtwrlIaesalkkIleleaeanaIrfcaahpbgikIvklrlrIshkelqreIsaf 763
0Y 770 TNDYNLAODLYLASSCPALALEMRDLOHWDsALQlAK-----HL 809
Db 764 ygefefeaeaklyldadgrldlatelmtlctdfrvqvqlylmgsgvsdqmeIawreIghnf 823
0Y 810 APDQIFRSIKRYALQLEFAGDYNVALAHYKEGITTGDNKEHDEACLAGVAMQSRMGDIRR 869
Db 824 anltswareareyukshlylegymealyhLe-----qIddleIk 860
0Y 870 GVNQALMKHPISVULKRDGALIEENNKQPSSEAOLYEGKLYDYKAASV--YIRSKNNAKGD 927
Db 861 cv-eelIpekspIrrl-laemaIasgmcsvevgallr--lgddqaavaatcvnltrqgeave 916
0Y 928 LLPHVSSPKIHLQYAKAKE--ADGRYKEAVVAVYENAKQOMQSVTRITYIDLHNNPKAVNI 984
Db 917 laqfrtqlprpqvtllakhaaqllqegrltkealIemqna----- 953
0Y 985 VREHOSIDGAKMARFPLQGLDYGSAILOPLVMSKCNNEAFTLAQOHKMEIYADITGSBD 1044

```

Db	954	--grhladaarllsq--maeregekraplrlrkkiylvlaallaehnk-----ava	999
Qy	1045	TTNEDYOSIALYFEGEKRYLQAGKEFLLCGGYSRALKHFLKCPSSDVAIEMAIETVGQ	1104
Db	1000	tleidyas-----grntlI-----delaIe-----	1019
Qy	1105	AKDELLTQMLDLHLLGENDCMKPKAKYFLRLYMLKQYREAAQVIAIITAREQSNAGYRN	1164
Db	1020	-----daaaIerIwhcaeyh-----fmllaqqlrlfyg----	1047
Qy	1165	AHDVLFMSYVIELKSKIKIRIPSEMANIIMLITLSYLLVKKIHVKNQDMKCARLLIRIVANNIS	1224
Db	1048	-----IvhsavvtavrlltdyevdIv-----	1067
Qy	1225	KFPSHIVPLTIVSTYIECHRAGLKNSAFSEFAMLMKPRERSKID-----AKYKKIEGWRR	1280
Db	1068	--pehlyallaaascdrafgtksaf-----mklegqahllpeatlqryelaaglfak	1119
Qy	1281	PDISIEEATTPPCPCCKLRLBECCELLCPGCKNSIPYCIATGRHMLK---DDW--TVCPHC	1335
Db	1120	ydpeditqtdrvdcyscgypvpdpsspcpcenarfpcacissgkpltpgpnmiwtictchhc	1179
Qy	1336	DEFPALYSEKIMLNRESTCPMC	1357
Db	1180	aapt-----eIsrhtcplc	1194
RESULT 10			
ID	AAU33013		
XX	AAU33013 standard; Protein: 72 AA.		
AC	AAU33013;		
XX			
DT	18-DEC-2001 (first entry)		
XX			
DE	Novel human secreted protein #3504.		
XX			
KW	Human; vaccination; gene therapy; nutritional supplement;		
KM	stem cell proliferation; haematopoiesis; nerve tissue regeneration;		
KW	immune suppression; immune stimulation; anti-inflammatory; leukaemia.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200179449-A2.		
XX			
PD	25-OCT-2001.		
XX			
PF	16-APR-2001; 2001WO-US08656.		
XX			
PR	18-APR-2000; 2000US-0552929.		
PR	26-JAN-2001; 2001US-0770160.		
XX			
PPA	(HYSE-) HYSEQ INC.		
PI			
PI	Tang YT, Liu C, Drmanac RT;		
DR	WPI: 2001-611725/70.		
XX			
PT	Nucleic acids encoding a range of human polypeptides, useful in genetic		
XX	vaccination, testing and therapy -		
PS			
XX	Claim 20; Page 698; 765p; English.		
CC	The invention relates to novel human secreted polypeptides. The		
CC	polypeptides and antibodies to the polypeptides are useful for		
CC	determining the presence of or predisposition to a disease associated		
CC	with altered levels of polypeptide. The polypeptides are also useful for		
CC	identifying agents (agonists and antagonists) that bind to them. Cells		
CC	expressing the proteins are useful for identifying a therapeutic agent		
CC	for use in treatment of a pathology related to aberrant expression or		
CC	physiological interactions of the polypeptide. Vectors comprising		
CC	the nucleic acids encoding the polypeptides and cells genetically		

CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AA029510-AA03304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 72 AA:

Query Match 3.8%; Score 274.5; DB 22; Length 72;
Best Local Similarity 85.7%; Pred. No. 7e-17;
Matches 54; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 274 IVCYMWGDRIMGFCGHPVYSTHTGELGEIFQARHKNDLSTIANVSQTLNKATC 333
Db 9 ILCYR-YgdgrImgfscghxvsthtgclgclqarhkdltlsavsqtlkvatc 67

QY 334 GDN 336
Db 68 em 70

RESULT 11
AAM79062
ID AAM79062 standard; Protein; 1241 AA.
XX
AC AAM79062;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1724.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PE 05-FEB-2001; 2001MO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
DR N-PSDB: AAK52195.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 4060-4063; 6221pp: English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibit activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1241 AA:

Query Match 3.6%; Score 263; DB 22; Length 1241;
Best Local Similarity 18.2%; Pred. No. 1.4e-13;
Matches 245; Conservative 197; Mismatches 506; Indels 396; Gaps 54;

QY 116 SSCYIWMDFNTNKTSDLDNGMDMSFLMSKVGSFLAVGTGKGLXLYNHQTSRKIPV 174
Db 113 ssdflglspeqkvsyk--hkssskllccswndgqyalgmfnqlstlnkgeekvkke 170

QY 175 -LGKHTKRITCGW-----NAENLXALGGEEDKMF-----TVSNOEG 209
Db 171 rpggsislplwscmpstwestfmmarenda---edvlnrylqelprstlksaaysqg 227

QY 210 DTRQTOVRSEPKNMQFLKKMDRTSAESMISV-LGKKTLFF-LNLERNPADLEF 267
Db 228 sgeeeeeeedspr-----ddnleerndllavdwgqkvsfyqsgkqlgkralnf 281

QY 268 QDPEGNIVCYMWGDR-IMGFCGHPVYSTHTGELGEIFQARHKNDLSTIANVSQT 326
Db 282 dp-----ccisyltkgelyllg--gsdkqvsflf-----kqgv----- 312

QY 327 LNKVATCGDMCIRIODVLDKMDIVLNLDEENKGLTSLMTDGOALASTORGSLHVF 386
Db 313 --flgtvg-----eqnswvltcgakrdsnoyvvugcdgtisfy 348

QY 387 LTKLPILGDACSTRIRAVLTSLLEYTVANRYEGELPTVSVDYERFVAVGLYHLANGMN 446
Db 349 qlfstvhgylkdrayrdsmtclvghllt--eqkvrlkckelvjkklaitylnlaqlpe 407

QY 447 RAMEYVLGENAVKLLKMDMETGTVASICLSHDAALFEGKVOULHLESETLDAQER-E 505
Db 408 klllyels---edlsdmly-----rvkekllkfecnllvvaanhllocekrld 455

QY 506 TRLEPAVDK-----CRILCHALTSDFLYGTDTGVVOYFYLEDMOFVNDYRNPV 555
Db 456 clstfsgkewqmeslliryklvlgppregllvgllkngqllklfdnlfalvllkqat 515

QY 556 SVKKIFPPDPNGTRLVFLDEKSDGFVPCPVNDATYEIRDFSTIKGVLMEMPRDKGVT- 614
Db 516 avrcldmsasrkklavvdenclclv---gidckellfegpnnsvawmncqcedmlcfsq 572

QY 615 -AYDDDKVYTVVFHKTIDGCAKVTLAGSTKVPFRHKRLLLYNG-ELTCQPGQSVNNIYL 672
Db 573 ggyylnlkastlfrvnrqlqg---fvvg-----yngsklfc----- 604

QY 673 STHGFLSLNKDXGDEFLRPMILANLMLKRFSDAMEMCRILINDEAANMLARACLNHMEVE 732
Db 605 -lhvf--sisavevpgsapmuy-ldtkllkeayqalcjvtdvdwclmealegldfe 660

QY 733 FAIRVYRRIG--VGIVMSLEBQK--GIEDYVNLAGHLAMFTNDVYLAODVYLASSCPT 787
Db 661 takkafirvgdlylqellssieerkkrgetnndlladvfsyqgklfneaklykrsgen 720

QY 788 AALEMRDLOHMSALDGLANHLARPDQIRFISKEVAIDLEFAGDVVNLALAYEKGITGDK 847
Db 721 lalemytdl-----cmfeyakdflyg-----sgdpx 745

QY 848 EHDEACLAGVAOMSIRMGDIRGVNOLKHPRSVYLKRDCALLENMKOFSEAOALYEGKL 907

```

Db 746 e-----tkmltkqadwar-----nlikepkaavemyisag 775
QY 908 YKDKAASYIRSKNMKAGDGLPHVSSPKIHLOYAKAKADGKRYKAAVAYENAKMOSY 967
Db 776 ehkalei-cgdhgyv---dm1ldl-----arkidaerepl1-----1 810
QY 968 IRIYDLNPNPEKAVNIVRETOSLDGAKMVAPEFLOLDGYGSAIQPLVMSKCNNEAFTLA 1027
Db 811 catylkkldspgya-----aetylkmgdlksivqlhvetqtwdeafalg 854
QY 1028 QQHNMKEITADITIGSEDTTNEYQSTALYFEGEKRYLQAGKFFLLCGQYSRALKHFLKCP 1087
Db 855 ekhpelk-----ddlympyagwlaendrfeeqakthkagrqreav----- 895
QY 1088 SSEDNVAIMEAFTVGQANDELLTNOLIDLGENGMKDKAYLRMLKQYREAO 1147
Db 896 -----qvleqltnavaesrf-----ndaayy--wmlsmqcldiag 930
QY 1148 TAIIRBESQAGNYRNADHVFESMYAEELKSOKIKIPSEMATNMLHSYILVKKHVKNG 1207
Db 931 dp---aqkdtmkgfkh-----fgriael-----yhgylahhrhtedp 965
QY 1208 DHMKGBMLIRVANNISKPSHVPILITSTVIECHR---AGLKNSAFSAAMLMPREYS 1264
Db 966 fsvhpelk---nistrllhslpkrtpsglskvklltflakqskalyarlahaydk 1021
QY 1265 ---KIDAKYKKIE---GMVRRPDISEIEEATTPCPFCFKLLPEELLCPGCKNSIPYC 1317
Db 1022 lqglylparfgskstlegtltlrkphdsee-----lvplc-----yr 1059
QY 1318 IATGRMLKDMVTWCPHCPALYS-----ELKIMLNTSESTCM 1356
Db 1060 cstmpnlmhlngvncincpgrffissasydvhlhvefylegltdeaa1ldlevlrpk 1119
QY 1357 CSFRLAAOLKRTSDCTOYLREE 1380
Db 1120 rddr-----qlelannssqllrlve 1139

RESULT 12
AAB43197
ID AAB43197 standard; Protein; 1252 AA.
AC AAB43197;
DT 08-FEB-2001 (first entry)
DE Human ORFX ORF2961 polypeptide sequence SEQ ID NO:5922.
XX
XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
XX vulerary; antipariatic; antiparkinsonian; noctropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritis; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antinflammatory;
XX antiviral; antibacterial; antifungal; antineumatic; antihypoid;
XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antinflammatory disease; coagulation;
XX thrombosis; contraceptive.
XX
OS Homo sapiens.
PN WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.

```

```

PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CUBA-) CUBAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI: 2000-602362/57.
XX N-PSDB; AAC77406.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 5103-5106; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytosolic; hepatotropic; vulerary;
XX antipariatic; antiparkinsonian; noctropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritis; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
XX antihypoid; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 1252 AA;

Query Match 3.6%; Score 260.5; DB 21; Length 1252;
Best Local Similarity 18.2%; Pred. No. 2.3e-13;
Matches 245; Conservative 197; Mismatches 507; Indels 395; Gaps 54;

QY 116 SSCITLMOANTKTKTSOLDNGMRDQMSFLMSKVSGSFANGTYKGNLXIYHNOTSKRIPV- 174
Db 71 ssdfglwspqekvsak--hkssskllccswtndgqylagmfnlglslnkngeekvk 128
QY 175 -LGKHTKTRTCGCM-----NAENLXALGEDKKMI-----TVSNOEG 209
Db 129 rpggsfslpwsicwpsrswesfmmnreda---edvlnvnylqelstlskaysysgq 185
QY 210 DTRQTVQVRSPEXNMQFLMKMDRTSAESKISYV-LGKTLFF-LNLNPDNDPADLEF 267
Db 186 seaeesepeeeddpr-----ddnleerndllavadgqkvsfyglsgkylgkralnlf 239
QY 266 QQDFGNYVMNYGGR-IMIGFSGHFVYSTHNGELGQELFQARNKNDNTSLAVSQT 326
Db 240 dp-----ccisylftkgey11lg--gsdkqvslft-----kdyv----- 270
QY 327 LNKVATCGDNCKIKIDVLDMKMYVILNLDEBNKGGTLTWTDDQLALASTORGSLEHF 386
Db 271 --rlgtvg-----egnswwtcqakpdsnyvvvgcgdlstsfy 306
QY 387 LTKLPILGDACSTRITAVLTSLLEVAVNPBEGELPITVSVDEPNFVAVGLYHLAVGMNN 446
Db 307 qlifstvnqlykdraydsmtdivgnhlt-egkvrltkckelvkklaiyrnlaiqlpe 365
QY 447 RAEFVVLGENAVKKLKDMEYLGCTVASTICLSHSDYAAALFEGKQVOLLISEILLDAQBER-E 505

```

```

Db 366 klllyels---edlsdmhy-----rvkekllkfeonl-lvvcanhllcgekr1q 413
OY 506 TRLEPRAVDK-----CRILCHALTSDFLIYGTDRGVGVQFYIEDMQFVNDVRHNP 555
Db 414 clsfsvkeewqmealllyikvlgppgrregllvlgknqgllkifvdlfaivllkqat 473
OY 556 SVKKIRPPDNGRLVPEIDKSDGFVCPVNDATVEIPDESPRIKGLVWENWMDKGVFI- 614
Db 474 avcldmasarkklavdendctlyv---didkellfgepansvawmtqcdmclcfsg 530
OY 615 -AYDDKVVTVYFHKDTIGAKVILAGSTKVRPAHKRPLLLYNG-ELTCTOSGKVNNTYL 672
Db 531 gyylnlkaactfvyhrqklqg---fvvg-----yngsklfc----- 562
OY 673 STFGFLSNLKDGXDELPRDLAHNLMLKRPDAMEMCRILINDAANNELARACTLHMEVE 732
Db 563 -lhvf--slsavevpsapmyqy-lqrklfkeayqlacjgvtdctdreamealegldfe 618
OY 733 FAIRVYRRIGN---VGIVNSLEQIK--GIEDVNLGLHMLAMFTNDYNLQDLYLASSCPT 787
Db 619 takkatfiryqdllylellsleerkkrgelndllfadyfsyqgkfheaaaklykrsghen 678
OY 788 AALEMRDLOHMDSALQLAKHLAPDOIPTISKRYAIQLEPAGDYVVALAHYEKGITGDNK 847
Db 679 lalemtydl-----cmfeyakdfllg-----sgdpr 703
OY 848 EHDEACLAGVAQMSIRMGDIRRGVNOALKNHPSVLKRDGAILENKKEOSEALQLEKGL 907
Db 704 e-----tkmltkqadar-----nlkepkaaaemylsag 733
OY 908 YYDKAASVYIRSKNMAKVGDLPRHVSPPRIHLOYAKAKENDGRYKAVVAUYENAKOMOSV 967
Db 734 envkatei-cgdhgwv---dmldi-----arkdkaerepll-----l 768
OY 968 IRIYDLHLNPEKAVNIVRETOSLGAKNVAFRLQLDGYSAIQFLVNSKKNNEAFITLA 1027
Db 769 catylklldspgya-----aeyllkmgdlkslyvqlhveqrvwdeaafalg 812
OY 1028 OOHNKKEITADIGSEDTNEDYQSIALYFEGSEKRYLQAGKFFLGGVSRALKHFLKCP 1087
Db 813 ekipeik-----dillympyqvlaendrfeeaqkafhkagrqreav----- 853
OY 1088 SSEDNVAIEMAIETVGOAKDELITNQLIDHLGENDGMPKDAKYLFRLYVALKQYBEAQ 1147
Db 854 -----qvleqlunavaesrf-----ndaayy--wmlsmqclidiag 888
OY 1148 TAIIRREOSAGNYRNANDVLFSMYAEKLSQKIKIPSEKATVLMILHSTILYKIHVKG 1207
Db 889 ad--peagctlmjgkfyh-----fgtrialael-----yhgylahlrhtedp 924
OY 1208 DHKKGARMLIRVANNSIKRPSHIVPLITSVIECHR--AGLKNSAFSPFAMLMRPBYRS 1264
Db 925 fsvhrpccflf---nlsrllhsipkdcpsgjskvkllflaqsalagryrlarhaydk 960
OY 1265 ----KIDAKYKKRIE--GWNRRPDISLEIEATTPCPCFLLRPECLELLCPGCKNSIPYC 1317
Db 981 lrglyparfqtgkeltelgtlitrakprhdeee-----lpllc-----yr 1018
OY 1318 IATGRHMLKDDMTVCPRCHDFPALYS-----ELKIMLTTESTCPM 1356
Db 1019 csefnpllnlmgvncncrpfllfssasydvlnhvefyleegltdeeaalsldlevlrpk 1078
OY 1357 CSERLNAOALKRISDCTOYLRTFEE 1380
Db 1079 rddr-----qlelannssqllrfive 1098

```

RESULT 13

AAB42248

ID AAB42248 standard; Protein; 1690 AA.

XX AAB42248;

XX

DT 08-FEB-2001 (first entry)

DE Human ORFX OREF2012 polypeptide sequence SEQ ID NO:4024.

XX Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;

XX vulnerability; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;

XX immunostimulant; osteopathic; antiarthritic; immunosuppressant; cardiant;

XX hypotensive; dermatological; immunosuppressive; antidiabetic;

XX antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;

XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;

XX neurodegenerative disorder; osteoarthritis; graft vs host disease;

XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

XX cholesterol ester storage; systemic lupus erythematosus; infection;

XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

XX bone damage; cartilage damage; antiinflammatory disease; coagulation;

XX thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shinketsu RA, Leach M;

XX WPI: 2000-602362/57.

XX N-PSDB; AAC76457.

XX Novel nucleic acids and peptides derived from open reading frame X,

XX useful for treating e.g. cancers, proliferative disorders,

XX neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 3208-3212; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

XX which represent the human ORFX open reading frames 1 to 3161. The ORFX

XX sequences have activities such as: cytostatic; hepatotropic; vulnerability;

XX antiproliferative; antiparkinsonian; neurotropic; neuroprotective;

XX osteopathic; antidiabetic; antiparkinsonian; neurotropic; neuroprotective;

XX immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;

XX antidiabetic; hypotensive; dermatological; immunosuppressive;

XX antiinflammatory; antibacterial; antiviral; antifungal; antineoplastic;

Sequence 1690 AA;

Query Match 3.6%; Score 259.5; DB 21; Length 1690;

Best Local Similarity 19.0%; Pred. No. 4.9e-13;

Matches 261; Conservative 187; Mismatches 463; Indels 463; Gaps 63;

0Y	76	YVWKIFEDHNGOKRSEIN---	LRGNVAMDMQKDODVLAIVAEKSSCIYLDAMTANKISOL	132	
0Y	77	YVWKIFEDHNGOKRSEIN---	LRGNVAMDMQKDODVLAIVAEKSSCIYLDAMTANKISOL	132	
Db	25	YVWKIFEDHNGOKRSEIN---	LRGNVAMDMQKDODVLAIVAEKSSCIYLDAMTANKISOL	132	
0Y	133	DNGMDHOSFLLMSVSGSFALVGTWKGMLXHYNOTSKRIEVLK---	HNKRIICGGMNA	189	
Db	83	YTESYVSVLITNCSOYGLISGHNADGILVYFFIDDEGSGSGKLIYNHOPPALAW	A	1400	
0Y	190	ENLXALGCEBKMTVSNQEGOTIROTOYRSEPRXNMQFELMKMDRTSAEBSM	-TSVVLG	247	
Db	141	TNSIVAAGCDKILVYAGKEGIMLQTFIDYSIRPGEREF-----	TLAVSPGSGSVVLG	192	
0Y	248	-----	KRTFLINLEPNRPMDLEFODFGNVCYNNWGGC-RIMIGSCGHE	294	
Db	193	SYDITRINWIPRISIW-----	EEAKKEI--INLYITELAWKIDSGISLWGLICG--	242	
0Y	295	VWISTHGELOEIPQA-----	RNKHDLTSLAVSOTLKNVATCGDNCI-----	KIO 341	
Db	243	-----	YQEGDCCLRSIYKIKFELVYGPBGVIVKHLNASSGTRVLKSHUYEVE	292	
0Y	342	DLVLD--KDMYVILNDEENKGLGTLSTWDDQLALSTQKSLHVLFLKLPILGDACSTR	4000		
Db	293	EKKILGKEYIVAHISE-----	LI	320	
0Y	401	IAYLTSLEEVYANVEEELRITVSDVERPNEVAVGLYHLNAGMNNRAME-----	YV	452	
Db	321	ISEIA-----	WGS	347	
0Y	453	IGENAVKALKDMEYIGTVASICLSHSDVAAALFEGKVOLHLESILDAOE---	ERETRLF	509	
Db	348	AGELIVGINDGLSVRYLEFEMP-----	HLISVINNERCQRTEDANKKIA	394	
0Y	510	PAVDKCRILCHALSDPLIKGTDRGVVUYEIEDMOFVNDYRHRPVSVKKIFRPNCTR	L	569	
Db	395	YLDIKTALAYD-----	LISGYNIGTVSHSRYDWELENE	434	
0Y	570	VFIDEKSDGEYVCPVNDATYETIPRDSPT--	IKGYLWENRPMDKGVTAJD-----	617	
Db	435	IFRDK-----	IRLHYLDESOSKFMILNFCISYNGWPVSGDVLVAGDINSJCWYLI	486	
0Y	618	---DDVVUYTVYFHKDPI--	OGAK--VILASIKRVFPAHKRPLLYNGELTCOGQSGKVN	668	
Db	487	EAPETVWFLTRGVDYIGLERGGKTEYMMMGVTVTEG---	LDGELIJEFTALDDN	541	
0Y	669	NIYVSTHGFSLNKDXGDELDRLPMILANMLKRESDAMEKCRILINDEAANNELARACLNH	728		
Db	542	---YIATAFLETL-----	EMTP-----	ETEAANKLISKALAE	572
0Y	729	MEVERALINVRKIGVWGIYVMSLEQIKGI-----	EDVYLLAGHLAMPTNOYNLA	776	
Db	573	QJLHIAECFSALIGVAKARFLHENEIADVSREYSGEGTDFYQVAVARILAMELEKNYKA	632		
0Y	777	QDLVLAASCPALAEEMRKDLOHMSDALOLNHLNLRDDI	PFISKEUYAIOL-----	825	
Db	633	EMIFLEGNAVEEOMGYELHTRDECLAVAEAKHRALEKIRTSUYGWLNDTQGEAGE	692		
0Y	826	---EFRGDUVNALAHYEKSTITDNKKEHDEACLAGVAAQMSIRMGDIRGVNOALKNHPSVULK	883		
Db	693	IGESGDDGLAASISYLYK-----	AGIPAKAARL-----	VLT 722	
0Y	884	RDCGAILENMKOFESPAADUYEKGUYLDOKAASVYIRSKNNMAKVGBDLPHVSSPKRILHO--	940		
Db	723	RE---ELLANTELVEHITAALKIGELYETA-----	QGLFEKIMPPQALAEYR	767	
0Y	941	---YAKARE--ADGRYKAEVAVAYENA-----	KOMOSVIRIYLDHLNPREKAVNIVRE	987	
Db	768	KGNAFMKAVELARIATAFRYEVVKLEAAGSDHLVGGKJDAAINHYIE-----	ARCSIKA	820	
0Y	988	TOSLIDGAM--VAREFLDGBGYSAIOFLVMSKCNNAEFTLAOON-----	NKMEIYADITG	104	
Db	821	IEAALGAYQWKAUYLDLQDINLASYK-----	YRVQADHYVASIYEYELAEELYT	870	

OY	1042	SEETNEDYOSIALYFEGSEKRYLOAGKFFLLC-----GOYSRLK	1081
Db	871	kgdrltkd---aIdmYlqa-grweqanlIamkcmrpedvsvlyltqagemakgkyreaser	926
OY	1082	HFLKCPSESDVWALEM-----ALEPVGQAKDELTLNOLIDHL-LG---ENGGMKRD	1128
Db	927	lylvcv--qepplaltumykhhklyldmrlrlygkhkhpdlssd---thlhlgkelaegrlqde	981
OY	1129	AKYLE-----RLYMLKOYREAAQTAIIIAEEOSSAGNYRNADVLFSSMYAEIKS	1178
Db	982	aeyhyleagewakatlumyryasglweeayr---vartqgga---nahkhaylwak---	1030
OY	1179	OKIRIPSMATNMLTILHSYLIVKILHVKNQDHMKGAMLRLRVANNISKFPSHIVPLTSTV	1238
Db	1031	-----sJlgeaaavrlllnklyg-----llaea	1050
OY	1239	TECHRAGIKNSAFSPAAMLMRPEYRSK---TDAKVYKKITGCAWRPDDISEIEEA	1289
Db	1051	vd-haa--dnscsfefafelrslalkhktpevhlykymfle-----degkfeea	1095
RESULT 14			
ID	AAE01339	standard; protein; 993 AA.	
XX	AAE01339;		
XX	17-JUL-2001	(first entry)	
XX	Human gene 24	encoded secreted protein fragment, SEQ ID NO:204.	
DE	Human; secreted protein; proliferative disorder; cancer; tumour;		
KM	focal abnormality; developmental abnormality; haematopoietic disorder;		
KM	immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;		
KM	inflammation; allergy; neurological disorder; Alzheimer's disease;		
KM	Parkinson's disease; cognitive disorder; schizophrenia; asthma;		
KM	skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;		
KM	cardiovascular disorder; angio-genic disorder; kidney disorder;		
KM	gastrointestinal disorder; pregnancy-related disorder;		
KM	endocrine disorder; infection; wound healing; vulnery;		
KM	cell culture; chemotaxis; food additive; gene therapy;		
KM	binding partner identification.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	WO200134769-A2.		
PN	17-MAY-2001.		
PD	01-NOV-2000; 2000MO-US30040.		
PE	05-NOV-1999; 9905-0163580.		
XX	30-JUN-2000; 2000US-0215130.		
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX	PA		
XX	Ruben SM, Komatsoulis GA, Wei P, Baker KP, Fiscella M;		
PI	WPI; 2001-308781/32.		
DR	New isolated nucleic acid molecule encoding a human secreted protein is		
XX	used in preventing, treating or ameliorating a medical condition -		
PT	disclosure; Page 51; 51pp; English.		
XX	AD05121-AD05203 represent cDNAs corresponding to 24 human secreted		
CC	protein genes, and AAE01332-AAE01311 represent the proteins they encode.		
CC	AAE01312-AAE01340 represent human secreted protein variants or fragments.		
CC	The secreted proteins and their genes are useful for preventing, gene		
CC	treating or ameliorating medical conditions, e.g., by protein or gene		
CC	therapy. Pathological conditions can be diagnosed by determining the		
CC	amount of the new protein in a sample or by determining the presence of		
CC	mutations in the new genes. Specific uses are described for each of the		

CC 24 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein fragment referred to in the disclosure of the invention.
 CC
 CC Sequence 993 AA:

Query Match 3.2%: Score 235.5; DB 22; Length 993;
 Best Local Similarity 18.8%: Pred. No. 2.8e-11;

Matches 218; Conservative 189; Mismatches 437; Indels 313; Gaps 52;

QY 97 CVAAMDNDKDDVLAV-IAEKSSCIYLMNDANTNKTSLD-----NGMRDQ 139
 -DB 58 cva--yakdgkrfagsads--vllw-----tsklegllkythndaigcveynptlhq 107
 QY 140 MS-----FLMKSQVSFLAVGVKGNLXLYNQTSRKIPVLGKHKRITCGGMENLX 193
 -DB 108 laecssaditglws-----peqkvsxkhsksk-----lccswtndgqyl 147
 QY 194 ALGCEPMKITVSNOEGDTIRQTOVRSEPMQPFLLMKMDRTSAESMISVGLGKTLFF 253
 -DB 148 algmfmglislrkngceekkerpgsjsplwsicmnspreendilavdagqkvsfy 207
 QY 254 -LNLNPNPNADLEFOODFCNIVCYNMVGDGR-IMIGFCGHEFVISTHTGELGOETFOA 311
 -DB 208 qlsagkqlgkdrainfdp-----ccisayfkgeyillg--gsdkqyslfc----- 249
 QY 312 RNNKDNLTSAVSQTLNKVATCGDNCIKIODLVLDKDMYIILDEBNKGLGLSWTDG 371
 -DB 250 ---kdgv-----rlgtvg-----eqnswwtgcgskpds 274
 QY 372 QLLALSTQNGSLHVFLLKLPLIGDACSTRATVYLSLEVTVANPVEGELITVSVDVEPN 431
 -DB 275 nyvvvgcqdgtisfyqlifstvhnglykdrayadsmcdvltvqhllt-ekkvrikckeyk 333
 QY 432 FVAVGYHLAVGNMNRAMPVVLGENAVKKLDMEYLGTVASICLHSDYAALFEKGVOHL 491
 -DB 334 klaiynrlaiaqlpekliylays-----edlsdmhy-----rvkeklikkfecl-1v 381
 QY 492 LISSEILDQBER-ETRLPAVVDK-----CRILCHALTSPDLITGTGTGVQYF 540
 -DB 382 vcanhlllcekrllcqlstisfgvkerewqmesllrylkivgpprgregllvgllngqllk 441
 QY 541 YIEDMDFVNDYRHPVSVKRTFFPDNGTRLVFEIDKSDGFYFCVUNATYETIRPESPFIK 600
 -DB 442 fvdnlfaivllkqatavrcldmasatrkklavndentclvy---didtckellfgepans 498
 QY 601 VLWENPMMDKGVFT--AYDDKVVYTVFHNKDTQGAQVILLAGSTKVPFAHKLPLLYNG-E 657
 -DB 499 vavnlqcedmlcfeggyyltikastfvrhqrklqg---fvvg-----yngsk 542
 QY 658 LTFQOTSGKVNNTYLSRHGSLNLKDXGPDELPRMLAHNMLKRFSDAMEMCRILNDEAA 717
 -DB 543 lfc-----lhvf--sisavevqgsapmyy-ldrkfkfeaygiacigvcltd 586

QY 718 WNELARACLHMEVEFAIRYRRIGN---VGIVMSLEQIK--GIEDYNLAGHLAMFTND 772
 -DB 587 wrleamealegldetkafkafirvgdlrylellssleerkkkggetndlfladvfygqk 646
 QY 773 YNLADLYLASSCPILAELRRDLOHWDLSALQAKHLAPDOIPTI-----SKEXA 822
 -DB 647 fheakrlkrsygeniallemtyldcmfeyekdfllgsqdketkmltlkqadvarnlkep 706
 QY 823 IQLEF---AGDYVNALAHYKGTGNGKNDKCLAGVADMSIRMGDIRGVQALKHS 879
 -DB 707 aavemyisagehvkate-----lcgch-----gwdmll--diarklkdaerepl 749
 QY 880 RVLKRDGCALTEEMKOFSEAAOLYEGKGLYDKAASYIRSKNMAK---VGDLLPHVSSPK 936
 -DB 750 ll-----cayllkklidspyyaaetyllmgdtkslvqhlvetqrdeafalgkpefktd 804
 QY 937 IHLQYKANEADGRYKEAVVAYENAKOMQSVIRIYLDHLNPNKAVNIYRETQSLDGAKM 996
 -DB 805 lymypaqwlaendrfeeaqafhkgqrteaevqy-1eqlln-----navaesrfndaa-- 856
 QY 997 VARFFIQLDYGSAIOLPLVMSKCNNAFTLAQOHNMK-ETVADITGSEDTTNDYQSIAL 1055
 -DB 857 -----yywmststgcltdiaqdpqkdmlgkly-----hfgrlae 891
 QY 1056 YFEGEKRYLQAGKFFLLCGYSRALKHFLLKCPSEEDNVATEMAIETVGAQKDELTN--Q 1113
 -DB 892 lylng-----yhaihr-----tedpfsh-----rpetlfnistr 920
 QY 1114 LIDHLLGENDGMPKD-----AKYLFRLYMALKQYREAAGTAIITAREOSAGNY---R 1163
 -DB 921 flhh-----slpkdcpysjkwkllfltl-----akgskaigayrlar 957
 QY 1164 NAMDVLFSMWAELEKSOK 1180
 -DB 958 haydkrlrglylparlqk 974

RESULT 15

AAB93308
 ID AAB93308 standard; Protein; 993 AA.

AA93308;

26-JUN-2001 (first entry)

Human protein sequence SEQ ID NO:12383.

Human; primer: detection; diagnosis; antisense therapy; gene therapy.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST:

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 MPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602
 full-length cDNAs defined in the specification, and for the detection
 and/or diagnosis of the abnormality of the proteins encoded by the

THIS PAGE BLANK (USPTO)


```

Db 247 KPSGSLIASTQDKP-----NOODIVFPEKNG-----L 273
Qy 163 IYNHOTSRKIPVLGKHTKRTTCGCMNAE-NLXALGEDKMITVSNQEGDITROTQVRSRP 221
Db 274 LHHFTF---LPEFL-KOEYKNDLMMNADSSVLAVERLED-----LORKESSIPRTCVQWLT 324
Qy 222 XNQOFLMKMDRTSAESMISVVLGKTLFPLNLNEPDNPADLEFQODGNTVCYMWG 281
Db 325 VGNVHMYLK-----QSLSFSTCGSKSLVSL-MMDPVLPYRLHLVLCQGMHVLAVDMWH 375
Qy 282 DGRIMGFSGHVFVISTHGE-----LCQELFQARBNK 315
Db 376 TTDRSVGDSSDLSNVAVIDGNRVLYTVFRQTVVPPMCTYOLLFPHRPNQVFLAHPOK 435
Qy 316 DNLTSIAVSQTLNKVAT--CGDNCKIKQDLY-----DLKDMYVILNLD 356
Db 436 SN--DLAVLDASNOISYKGD-CPSPADPVLKGVAGSGFKVCLRTPHLEKRTKIQFEN 482
Qy 357 EENKG---LGTLSMTDQGLLALS---TORGSLHVFLLKPLIGDACSTRIAVLTSL 407
Db 493 NEDQDVNPLKGLTLTMEEDVFLAVSHSPRSVTH-HLT-----AASSSEM----- 538
Qy 408 LEVTVANPVGELPIYVSVDYEPNPAVGLYHLAVGMNRPVYVLAGENAVKLLKDMETL 467
Db 539 ---DEHGQLNVSSSAVD----- 554
Qy 468 GTVASICLHS---DYAAALFEKVQLHLESEILDAOEERTRLP----- 510
Db 555 GVTISLCNKSRTSVVLQDLADGQIFKYLWSPSLAIKPKMNSGFPVRFPTQOTELAM 614
Qy 511 -----AVDKCR-----ILCHALTSDFLIYGTDTGVVQYVEIDMQF--- 547
Db 615 IGEBCVLGTLTDCRFINDIEVASNITSPFAVYDEFLLTTHSHTCQCFCLRDSSEKTLQ 674
Qy 548 ----VNDYRHPVSKKIFPNNGTRLYFIDEKSDGFVYCVNATYEIPDSFTIGVLM 603
Db 675 AGLSNHHVSGEVLKRV--ERGSRTVTV-VPODTKLYLQMPRGNELVNHRALVLAQI- 729
Qy 604 ENMPMDKGVFIAYDDKVVTVYVFNKDTIOGAKYILAGSTVPAFRHPLLYNGELTCORQ 663
Db 730 RKW-LDKLME-----KEAFECMRKLRLINLNI-YDHNPKVFLGVEFFINQ 773
Qy 664 SGKVNNTIYSTGFLSNLKDQDELRPM---LAHNLMLKPSDAMEKRIINDEAAMN 719
Db 774 IDSVNHNIL--FTELKE--EDVTKTMYPAVTVSSVYLSRDPDNKIDLYCDAMRAVA 827
Qy 720 ELA---RACL---HHM-----EVEFAL-RVYRIRGN-----VGIVMSLEDKIGIEDY 759
Db 828 ESTNPKYICLSITSHVKKTPPELETYVLOVHELQSNAPSDPDPAVSAEELKTLHLVDY 887
Qy 760 NLAGHLMETNDYNLADLYLASSCPIALALEMRRLQHMDSALQAKHLAPQIPEISK 819
Db 888 NELYDH-SLGTYDF---DLVL-----MVAEKQKDPKEVLPFLNLTKMETNYQFETID 937
Qy 820 EYAIQLEFAGDYVNALAHYEKGITGDNK--EHDEACLAGVAQMSIRMGDIRGVQALX 876
Db 938 KY-----LKRREKAIGHLSKCGPEYFPECLNLKIDKNL-----YNEALK 976
Qy 877 --HPSRYLKRDC---GALLENKQFSEAOLYKGLGYIDKASVYIRSNNAKAVDDL 930
Db 977 LVSPPSQOYODISIAVGEHLMQEHMTEPAGLMEARGCAHKAASLALTCGNW----- 1028
Qy 931 HVSSPKIHQYAKAKADAGYKEA--VVAYENAKQMSQSVIRIYLDLHNNPEKAVINRET 988
Db 1029 -----KQALCVAAQLNNTKQOLYGL-----GRTLAGKLYEOR 1060
Qy 989 QSLDGAQVYARFQLQDYGSAIQFLWMSKCNNEAFTLAQOHKMEI----- 1035
Db 1061 KHLIDAA-VYLEBCAQ--DYEAVALLLLEGAAWEEALRLVYKYWRDLIIETNVPSTLEAQ 1117
Qy 1036 --YADIGSEDTT-----NEDYQSIALYFEGEKRYLDAGKFFL----- 1071

```

```

Db 1118 KNYMAFLDSQATWATFSRKRKRLVRELKBOAQAGL--DDEVPHQGESDLFSETSSVSG 1175
Qy 1072 --LCGOYSRA-----LKHFLKPSCEDNVAAIEATVGO-----AKDE 1108
Db 1176 SEMSGKYSNSNRTISARSSKNRKAERKHSLEKESPLEDLALLESEVONTENLKD 1235
Qy 1109 LRTNOLIDHLGENDMPKDAKYLFRLYMALQYREAQO---TALITIAI----- 1154
Db 1236 -----VHHL-----KVLV-LFEFDEQGRLEQAFEDTLQLMERSLPEIWTLY 1278
Qy 1155 EEOAGNYRNADVLFMSYAEKLSOKIKIP 1184
Db 1279 QONSATPVLGPNSTANSIMASYOQKTSVP 1308

```

RESULT 2
US-09-286-891-2
Sequence 2, Application US/09286891
Patent No. 6172195
GENERAL INFORMATION:
APPLICANT: Cohen, Lucy
APPLICANT: Baeuerle, Patrick
TITLE OF INVENTION: IKAP Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/286,891
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/971,244
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1332 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-286-891-2

Query Match 2.3%; Score 167; DB 4; Length 1332;
Best Local Similarity 18.7%; Pred. No. 1.2e-05;
Matches 253; Conservative 186; Mismatches 455; Indels 456; Gaps 66;

```

Qy 51 TWGAPIQIPAMQKTSQNTYLA-----TGADYIVKIFEDRHG--QKRSIELNPGNCVAMDM 102
Db 199 TMRG-----DQAFVAVSVCPETGARK-VRYMNEFALQSTSE-PVAGLGPALAM 246
Qy 103 DKDGDVAVIARSSSCTIYLDANTNKTSQLDNGMRDQMSFLMSKVSFLAVGYKGNLX 162
Db 247 KPSGSLIASTQDKP-----NQODIVFPEKNG-----L 273
Qy 163 IYNHOTSRKIPVLGKHTKRTTCGCMNAE-NLXALGEDKMITVSNQEGDITROTQVRSRP 221

```

```

Db 274 LHGHT---LPL-LKDEVKVNDLLMNAOSSVLAVRLED-----LOREKSSIPKTCVOLMT 324
QY 222 XNQFPLMKMDRTSAESMISVVLGKTKTLEFLNLBNPONADLEQOQDGNIVCYNWYG 281
Db 325 VGNVHMYLK-----QSLSESTCGSKATVSL-MMDPVTYRHLVLCOGMHTLATIDMW 375
QY 282 DGRIMIGFCGHPVISTATGE-----LGQEFPOARNHK 315
Db 376 TTDRSVGDNSDLSNVAIDGNRVLTVPFQTVVPPMCTYOLLFPHRVQVFLAHPQK 435
QY 316 DNTSTAIVSOTLKKVAT--CGDNCIKIOLV-----DIKMYVILIND 356
Db 436 SN--DLAVLDASNQISVYKCGD--CPASADPTVKGAVGSGFKVCLRPHLEKRYKIOFEN 492
QY 357 EENKQ-----LGTLSMTDDGOLLALS---TORGSLHVELTKLPIIGDASTRIATVLSL 407
Db 493 NEDQDVPRLKGLLITVIEEVEFLAVSHSESPSPSVH-HLT-----AASSE----- 538
QY 408 LEVTANPVEGELPTVSVDPENFVAVGLYHLAVGMNBRMFYVLGENAVKIKLDMEYL 467
Db 539 -----DEERGLQNVSSAAVD----- 554
QY 468 GTVASICLHS---DYAALEFGKVQHLLESELDAQERETRLP----- 510
Db 555 GVIISLCNSKTSKVYLQDLADQIFKYLWESPSLAIKPMKNSGFPVRFPYCTQETELAM 614
QY 511 -----AVDDKCR-----ILCHALISDFLIYTDGVOYFIEQMOMF----- 547
Db 615 IGESECVLGLTDRKRFINDIEVASNTSPFAYVDFELTLTHSHTCQCFCLRASFTTLO 674
QY 548 -----VNDYRHPVSVKIFPPNGTRLVFIDEKSDGFYVCVNDATYIEPFPSTIKGLM 603
Db 675 AGLSNHHVSGEVLKRV---ERGSRIYTV-VPRDTKVLQMPRGNEVYVHRLAVLAQI- 729
QY 604 ENRPMDKGFIAVDKDVYTVVYVHKDTIQGAKVILAGSTKVPRANKRPLLNGELTCOTO 663
Db 730 RKM-LDKLMF-----KEAFECMRKLRLINLPI-YDHNPKVYLGAVETFIQO 773
QY 664 SGKVNINYLSTHGFSLKXGPDRLRPM---LAHNLMKRPSDAMEKRIILDEAMN 719
Db 774 IDSVNHNIL---FTLEKE--BDVTKMYRPAVYTSVYLSRDPGKIDVLCAMRAYM 827
QY 720 ELA---RACL---HHM---EVERAI-RVYRIRIGN-----VGIVMSLEOIKIEDY 759
Db 828 ESINPMKYCSILTSYHKKTPELETYLOKVHLEQGNAPSDPDAVSEALKYLLHLYDV 867
QY 760 NLLAGHLAMFTNDYNYLAODLYLASSCPIALALEKRRLDLOHWSLALOKHLAPPOIPISK 819
Db 888 NELYDH-SLGTYDF---DLVL-----MVAEKSOQDPKEYLPFLNLTLLKMETVYQREFTID 937
QY 820 EYAIQLEFAGDVYNALAHYKGTGDNK---EHDEACLAGVAMQSTIMGDIRRGVNOALK 876
Db 938 KY-----LKRYEKAIGHLSKCGPEYFPPECLNLIKDKNL-----YNELAK 976
QY 877 --HPSRYLKRDG---GALLENMKOFSEAAQLYKGLYDYKASVYIRSKNMAKVGDLRP 930
Db 977 LYSPPSSQOYODISIAVGENHLMQEHMWERPAGLMAFGCAHMKALSAFLTCGNW----- 1028
QY 931 HVSPPKIHLOYAKAKADEKRYKA--VVAVENAKQWOSVIRIYLDHLNPEKAVNIYRET 968
Db 1029 -----KQALCYAOLNFTKDLVGL-----GRTLAGKIVEOR 1060
QY 989 OSLDGAKMVARFFLOLDGYSAIOTPLVMSKNNEAFTLQOHNKMET----- 1035
Db 1061 KHLIDAA-MYLEBCAQ--DYEAVALLLLEGAMBEALRLVYKYNRDLIETNVKPSILEAO 1117
QY 1036 --YADIGSEDTY-----NEDYQSIALYFEEGKRYLQAGKFFL----- 1071
Db 1118 KNVMAFLDSTATFYSRHKRLLVVRELKEQAQAGL--DDEVYHGGQSDLEFSESSVYSG 1175
QY 1072 --LCGQYSRA-----LKHFLKCPSSSENVVAIEMALETVQO---AKDE 1108
Db 1176 SEMSGKYSHSNRISARSSKNRRKAEKHKSLKEGSPLEDLALLEALSEVQNTENIKDE 1235

```

```

QY 1109 LITNQLIDHLGENDGMKPAKYLFRLYMALKOYREAAO-----TATIIAR----- 1154
Db 1236 -----VYHIL-----KVLF-LFEPDEGRLOKAFEDTLQUMERSLEPIWTLLY 1278
QY 1155 EEOAGNYRNADVLEFSMVAELKSORIKIP 1184
Db 1279 QQNSATPVLGPNSTANSIMASYQQOKTSVP 1308

RESULT 3
US-08-775-009-35
; Sequence 35, Application US/08775009
; Patent No. 5935783
GENERAL INFORMATION:
; APPLICANT: Gong, Wellong
; APPLICANT: Emanuel, Beverly S.
; APPLICANT: Budarf, Marcia L.
; APPLICANT: Roe, Bruce
; TITLE OF INVENTION: No. 5935783el Genes Mapping in the Digeorge and
; TITLE OF INVENTION: Vellocardiofacial Syndrome Minimal Critical Region
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn, Kurtz MacKiewicz &
; ADDRESSEE: No. 5935783ris, LLP
; STREET: One Liberty Place - 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/775, 009
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yalko
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CH-0681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-34100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 871 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-775-009-35

Query Match 1.9%; Score 140.5; DB 2; Length 871;
Best Local Similarity 19.0%; Pred. No. 0.0015;
Matches 158; Conservative 113; Mismatches 257; Indels 305; Gaps 39;
QY 445 NNRAMPYVLGENAVKRLKMEYIGT--VASICLHSD--YAAALFE-GKVQHLIESEIL 498
Db 113 MALAKYIIDSNNNPERLRENPYYDSVAVGKYEKRPDLACAVAYERGQDLELI----- 167
QY 499 DAOEBERTLFPVADKCRILCHALTSDFLIYGTDTGVOYFIEQMOPV---NDYRHP 554
Db 168 --NVNENSLFKSL-----SRVLYRRKDP-----ELMGSVLLESNPYRRP 205
QY 555 VSVKRIEPPDNGRLRVFIDEKSDGFYVCVNDATYIELPDSPTIKGVLMENMP-----M 608
Db 206 L-----IDQ-----VVQTALESFQDPEEVSIVKAFMTRADLNLELTL 244
QY 609 DKGVFIAYDDKVVYTVVYVHKDTIQGAKVILAGSTYKVFPAHKPILLYNGLTCTOGSKVN 668

```

```

Db 245 EKIIVL-----DNSVFSERHILQNLITLTAIKADTRV-----MEYINRLD-NYDAPDIA 292
QY 669 NIYLSTHGF-----LSNLKDXGPDELRPMIAHMLMKRPSDAMEKRIILINDEAANIELAR 723
Db 293 NIAISNELFEBAFAIRKRDVNTSAVOVLEIHGNLDR---AYEFERENEPNVSOLAK 349
QY 724 ACLHMEVEFAIRVYRRIGNVIGVLSLEQIKGIEDYNLLAGHLAMFTNDYNLAQDLYLAS 783
Db 350 AOLQKGMVKEAI-----LYKDAQVASESKTLEABELLOQFLDE 796
QY 784 SCPIALALEMRDLO---HMSDA---LQLAKHLAPOIPRISKEVATOLEFAGDYVALAH 837
Db 368 DDPSSTMEYVQAANTSGNMEELVKYLQMARKKARES--YVETE---LLEFALAKTNRLAE 421
QY 838 YEKGLITGDNKEH-----DEACL-----AGVQMSIRMGDIRGVNOAL 875
Db 422 LEEFINGPNNAHIIQOVGDRCYDEKMTDAKLXNNVSNFGRSLASTVHLGELYQAVDGAR 481
QY 876 K-HPSRYLKRDCGAILENK-----QFSEAAQLYER 905
Db 482 KANSTRFKMEVCFACVDGKEFRLAQMGCHIVYHADELELINYQDRGFEEELITMLBA 541
QY 906 GLYVOKA-----ASYVIRSKN-----NAKVDDLPHV--SSPKIHL----- 939
Db 542 ALGLEBAHGMFTELAILYSKFRPKQMRHELEFMSRVN--IPKYLRAEQAHMAELVE 599
QY 940 QYAKAKKAD-----GRYKAVAVAYENAKOMOSVIRIY----- 971
Db 600 LVDKKEEYONAIITMMNHPTDAMKEGQFQDITIKVANVELYRRAIOFYDEFPILLNDL 659
QY 972 -----LHLLNPEKAVNIVRETQSLDGAKMYAR-----FLQLGQDYS 1009
Db 660 MYLSPRLDH-----TRAVNFSKYKQPLPVKPYLRQVQNNHNSVMSLNLEFITEEDY-Q 714
QY 1010 AIOFLVMSKCNNEAPFLAQOHNMKEIYADIIIGSEDTNEDYOSIALY-PEGKRYVQACK 1068
Db 715 AIRTSDAIDANDNISLAKREKHELI-----EFRIIAYLEFKGNKRMQSQVE 762
QY 1069 FFLIQGVSRALKHLKCPSSSDNVAIEMAIEITVGQAKDELITNLDHILHGE 1121
Db 763 ---LCKKDS-----LYKDAQVASESKTLEABELLOQFLDE 796

```

```

RESULT 4
US-08-874-266-2
: Sequence 2, Application US/08874266
: Patent No. 5955279
: GENERAL INFORMATION:
: APPLICANT: Gatti, Richard A.
: TITLE OF INVENTION: ATAXIA-TELANGIECTASIA: MUTATIONS IN THE ATM GENE
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobb, Martens, Olson and Bear
: STREET: 620 Newport Center Drive 16th floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: USA
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/874,266
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Ways Vensko, Nancy

```

```

: REGISTRATION NUMBER: 36,298
: REFERENCE/DOCKET NUMBER: DCIA006,006A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-235-8550
: TELEFAX: 619-235-0176
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3056 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
US-08-874-266-2

```

```

Query Match 1.9%; Score 136.5; DB 2; Length 3056;
Best Local Similarity 18.0%; Pred. No. 0.034;
Matches 236; Conservative 176; Mismatches 420; Indels 481; Gaps 62;

```

```

QY 322 VSQTLNKVATCGDNCKIKIO-----DLVLD-KDMYVIL-----NLDEENKGLTSLMTD 369
Db 455 VLRCLEVALCODKSNLESSQKSDLIKMKIKWICTTFRGISSEIOQAEISGL--LGALI 512
QY 370 DQOLLAISTQSLHVFILTKLPILIDACSTRIVYLSLEVTVANPVGELPITVSQVE 429
Db 513 QGSLVEVDREPKFLPTGSACRSPCAVCCLLALPTISI-----VGAVKMGIE 560
QY 430 PNFVAVGLYHLAVGNMNAFVLGEMNAVKL-----KDMYLTGVASICHSIYAAAL 483
Db 561 QNMCEV-----NNS--FSLKESIKWLEFYQLBQEDLNSTVEPPI-LHSNPHLY 607
QY 484 FEG-----KVQLHIES--EILDAQEERETRLPAVD-----DKRCILCHAL 523
Db 608 LEKILVSLMKKCKAMNFPQSVPECHQKKEBELSEVEBELQQTTFDK----- 659
QY 524 TSDFLITGDTGVQY-----FYIE-----DMQFVNDYRHPVSVKKIFDP 564
Db 660 -MDFLITVIREGCIERHOSIGFSVHQNLKESIDRCILGLSEQLLNYSSEIT----- 710
QY 565 NGRILVFLIDKSDGVVQPVNDATYEIDPFSPTIKGLWEMNPMKGVFIADVDDKYTY 624
Db 711 -----NSETLVRC-----SRLLVGL-----GCT-----CYMG 733
QY 625 VFHKDTIOGAKYTLAGSTKVPFAHKKPLLYNGELTCQTQSGKVNNIYLSHGFSLNLKDX 684
Db 734 VIAEEAVKSELFQKANSIMQCAGESITLFFKKTNEEFRIISLRMMQDCTRCLSNCKRK 793
QY 685 GPDELRP-MLAHNMLMKRPSDAMEKRIILINDEAANIELARACLHMEVEFAIRYRRIGN 743
Db 794 SPNKIASGFFRLILSKLMDIADICKSL-----ASFIKKPPDR--- 832
QY 744 VGIVMSLE-----QIKGIEDYNLLAGHLAMFTNDYNLAQDLYLASSCPILALEMRDLOH 798
Db 833 -GEVESMEDDTNGNLMEVEDOS---SMNLF-NDY-----PDSSVDANEIGE 874
QY 799 WDSALQAKHLAPDOIPRISKEVATOLEFAGDYVALAHYEKGITGDNKEHDEACL----- 854
Db 875 SGGTIGAINPLAE-----YLSKQDLFL-----DMLKFL-----CLCTYT 910
QY 885 AGVQMSIRMGDIRG-----VQALKHPSR-----VLKRDGCA-----ILENM 893
Db 911 AQTNTVSFRADIRKRLMLIDSSYLEPTKSLHLMYLMLKELPEGEFPLPMEDVLELL 970
QY 894 KQFSEAAQLEYEGLYDKAASYIRSKMKAKGDLPHV-----SSPKIHOYAKAKKAD 948
Db 971 KPLSNVCSLYR-----RDQVCKT--ILNHVLAHYKVLGQSNMDSNETRDAO 1015
QY 949 GRKRAVAVAYENAKOMOSVIRIYDLHLLNPEKAVNIVRETQSLDGAKMYARFFQLGQYG 1008
Db 1016 GQFLTVIGAF-----W-----HLTKERKIYISVM-----ALVNCILTLLEADYS 1056

```


RESULT 7
US-08-287-959-1

```

Db 764 EFRSRMFL--KKQIPATCISQQRGKQKKKAYQDRLAYLRSHKDEYVVKIQSLARMHQ 820
      ::||::||::||::||::||::||::||::||::||::||::||::||::||::||

```

```

0Y 1059 GEEKRLONGKEFELLCGOVSRALKHFLKCPSSDDNAIEMAEITVGAKD-ELTTNOGLDH 1117
Db 821 ARKRRLDQYFRDHINDIIRKIAFIIRANKARD-----YKTLINADPPVAVVRKFEVH 874
0Y 1118 LIGENDGPKRPAKYFRLRYMLKQYREAAQTAAIIAREEQSAGNYRNAHDVLFMSYAEIK 1177
Db 875 LLDSD---QD---FQEBELDMKKREVIITL---RNOOLENLNMDIXIGL---LV 921
0Y 1178 SOKIKIPSEMATNMLHSYLLVTKIHYKNGDHMKGARM 1216
Db 922 KNKITLQD-----VSHSKKLTK--KNKEQLSDMMMI 951

RESULT 8
US-08-508-836A-8
Sequence 8, Application US/08508836A
Patent No. 5777093
GENERAL INFORMATION:
APPLICANT: Shioh, Yosef
APPLICANT: Tagie, Danilo A.
APPLICANT: Collins, Francis S.
TITLE OF INVENTION: Ataxia-Telangiectasia Gene
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Relsing, Ethlington, Barnard & Perry
STREET: P.O. Box 4390
CITY: Troy
STATE: Michigan
COUNTRY: US
ZIP: 48098
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,836A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: P-313 (TAU)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 689-3500
TELEFAX: (810) 689-4071
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3056 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-508-836A-8

Query Match 1.8%; Score 131.5; DB 1; Length 3056;
Best Local Similarity 17.6%; Pred. No. 0.098;
Matches 231; Conservative 178; Mismatches 425; Indels 477; Gaps 60;

0Y 323 VSQITINKAVATGDCNCKIKQ-----DLVLDKDMYVILNL-----DEBNKGLTSLKSTDDG 371
Db 455 VLRCTEVALCODKSNLESSOKSDULKMNIMWCITFERGISSEOKQOAFNGLGAIIOG 514
0Y 372 OLLALSTORGSIAHVEFLTKPIILGDACSTPIAYLTSLEVTYVANPYEGELPIIVSYDVEPN 431
Db 515 SLVEVDREPKLFETSSACRPSCPAVCCCLTALTTST-----VPGAVKMGIEON 562
0Y 432 FVAAGVLYHLAVGMNRRAMEYVIGENAVKKL-----KDMEYLGTVASICLASDYAAALPE 485
Db 563 MCEV-----NRS--FSLKESIMKWLFLPYOLEGDLNSTEVPPI-LHSNPHVLVE 609
0Y 486 G-----KVQHLIES--EIIIDQEEERTLPEAVD-----DKCHILCHALTS 525

```

D	b	610	KILSLTKMKCKKAAMFQSVRCESEHHNKKDELSFSEYEBELFLQTTDDK-----M	660
Q	y	526	DFLLYGDTGVVY-----FYIE-----DMQVNDYRHPVSVKIFPPNG	566
D	b	661	DFLLYRECIEKHOSSIGFSVHQNLKESLDRLGLSLBOLLNYSSEIT-----	710
Q	y	567	TRLVFEKSDGVYVCPVNDATYELIPDESPITKGVLMEMWPMDDKGVFLAYDDKAYUYTVF	626
D	b	711	-----NSETLVRC-----SRLVGVY-----GCV-----CYMGVI	735
Q	y	627	HKDIOGAKYILGSTRVPAPKPLLYNGELTQSGQKVNITYLSTHGFSLNKDGP	686
D	b	736	AEEBAYVSELPQANSLMOCAGSITLFFNKTRNEERIGSLRNMQOLCTLSNCTKSP	795
Q	y	687	DELAP-MLAHNMLKRESDAMEMCRILNDEAAMNELARACLHMEVEBALRYRRIQNG	745
D	b	796	NKIASGFPLRLTSLKMLNDIADICKSL-----ASFTRKPFDR---G	833
Q	y	746	IYWSLE-----QIKGIEDYNYLAGHLAMTNDYNLAODLYLASSCPIALEMKRDLOHD	800
D	b	834	EVESEMEDTNGNIMEVEDOS-----SNMLP-NDY-----PDSVSDANPEGSQ	876
Q	y	801	SALDOLAKHLARDPIPSKRYATQIEFADYUNALAHYEKGTGONKEHDEACT---AG	856
D	b	877	STICAIPLLEE---YLSKODLFL---DMLKFL-----CLCYTTAQ	912
Q	y	857	VQMSIRMGDIRGQ---VQOALKHPSR-----VLKRDGCA-----ILENNKQ	895
D	b	913	TNTVSPFAADIRKRLMLIDSSTLEPTKSLJLHMVLMILKELRGEBEYPLRPEDVELLKP	972
Q	y	896	FSEAOLEYEGLYYDKAASYITRSKKMAKVGDDLPHV-----SPFKHLOYAKAKEADGR	950
D	b	973	LSNVCSLYR-----RDOOVCKT---ILNHVLHVNNKLOQSMDENTRDAQQ	1017
Q	y	951	YKEAVVYENAKOMQSVTRYLDLNNPKAVANYRETSQDGAMVAPRFLOGDGSA	1010
D	b	1018	FLTYIGAF-----W-----HLTKERRYFSYRM-----ALVNCIKTLLPADPSKWM	1058
Q	y	1011	IQPLWMSK---CNNEAFT---LQOHNKMEIYADIGSEDTNEDYOSIALYECEK---R	1062
D	b	1059	ALVWVKDKDPVNEVFPQFLADNHQVRLA-----ASINRLQDFQDGSSR	1106
Q	y	1063	YLOA-----GKFLLCGOYSALAKHFLKPSSEDN-----VAIDMAIET	1101
D	b	1107	LKLKALPLKLOOTAFENYALKAQGMKREMSISAENPETIDEIYNRKSVLLTILAVVLSCP	1166
Q	y	1102	VGO-----AKDBELLNOLIDHLG---NGMFKDAKYPLRYLMALK-1140	
D	b	1167	ICEQOALFALCKSVKEGDELPRLVKKVLEKSEVETGTYRLEDPFMSHLDYLVLEKLNQD	1226
Q	y	1141	-QYREAAQTAIIAREBOSAGNYRMAHD-----VLFSMYAEKLSOKIKIPSE	1186
D	b	1227	TEVLMSSFPILL-----NTNIEDPYRSQYKVLPHLVIRSHPEVANSIAHQIED	1278
Q	y	1187	MATNMLTLHSYILVKI-----HYKGDHMKGARML-----	1216
D	b	1279	WKSLLTDCPFKILVNIILPFAVEGTRDSGMAOQRETAQVUYMLSENLKQDHLFIS	1338
Q	y	1217	-----IRANNISKFPBHIYPIILSTVYIECHR	1243
D	b	1339	NLPEIIVELLMTLHEPANSSASOSTDLCDFSGDOLAPRPHFPBHVIKATPAATYSNCKH	1398
Q	y	1244	AGLEN-----SAFSPAALMR-PEYRSKIDAKYK---KTE-----	1275
D	b	1399	TKLKSILELTSKSPDSQKILLAIQGAQMETNNVYKHHILKIHYLFLPSSLLKDKSGLG	1458
Q	y	1276	---GMYRRPDISIEEATTPCPC-----KPLPECELLCGCKSNIPYCINATGRHML	1375
D	b	1459	GAMAFVLRDVIYTLIHINORPSCIMDVSLRSFSLC-CDLSQVOCQTAFTVC-----	1509
Q	y	1326	KDDVTGCPHCDPALYSELKIMLNTFESTCPMCSERLMAAQIKKIDCQOYL	1376
D	b	1510	KD-----ALENHVHVIQGT---LIPLYVEQVQV---KOVIDLLKYTL	1545

```

RESULT 9
US-08-629-001A-3
; Sequence 3, Application US/08629001A
; Patent No. 5858661
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
; TITLE OF INVENTION: GENOMIC ORGANIZATION
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 5858661thwestern Hwy.
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/629,001A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 2290.00032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 539-5050
; TELEFAX: (810) 539-5055
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-629-001A-3

Query Match 1.8%; Score 131.5; DB 2; Length 3056;
Best Local Similarity 17.6%; Pred. No. 0.098;
Matches 231; Conservative 178; Mismatches 425; Indels 477; Gaps 60;

DB 323 VSQTLNKVATCGDNCIKIO-----DIVLDKMYVILNL-----DEENKGLGTLSTWTDG 371
DB 455 YLRCTLEVALCDCKRSNLESOSKSDLLKLNKIMCTFRGISSEKQAEFGILGAIIOG 514
DB 372 QLLALSTOGSLHVFLLTKPIILGDCSTRIAYILSLLEVTVANPNVGCGLPIYSDVVEPN 431
DB 515 SLVEVDREKWKLFPTGSACRPSCPAVCCLTALTLTST-----VPGAVKMGIEGN 562
DB 432 FVAAGLYHLAVGNRRWFYVLGENAVKRL-----KDMYELGTVAASICIHSDYAAALFE 485
DB 563 MEV-----NRS--FSLKESIMKWLFLYQLEGDLNENSTEVPI-LHSNFRPLVLE 609
DB 486 G-----KVQLHLES--EILDQERETRLPRAVD-----DKCRILCHALTS 525
DB 610 KLVSLVLMKMKKAAMNFQSVPECEHHNKKKELSFSEVEELFLQTFPDK-----M 660
DB 526 DELIYGTOTGVQY-----FYIE-----DMQFVNDYHNPVSVKKIPDPNG 566
DB 661 DFLTVIREGIRKHOSSIGFSVHQLKESLDRCLLGLSEQLNNVSSSEIT----- 710
DB 567 TRLVTFIDEKSDGFYVCPVNDATYETIPDSPTIKGVLMENWMPDKGVFLAYDDDKVYTYVE 626
DB 711 -----NSETLVRC-----SRLVGVYL-----GCY-----CYMGVI 735

```

```

DB 627 HKDTIGAKVILLAGSTKVPFAKPLLNGELTCQTSKGVNNIYLSTHGFSLNKKDXCP 686
DB 736 ABEBAKSESLFQKANSIMWCAGESTITLFRKKTNEEFRISSLNMQGLTRCSNCTKSKSP 795
DB 687 DELRP-MLAHNIMLKRFSPAMKCRILNDEAAMNELARCLHMEVEFAIRYRRIGANG 745
DB 796 NKIASGFLLRLTSLKIMNDIADICKSL-----ASFTRKPFDR-----G 833
DB 746 IYMSLE-----QIKGIEDYNLAGHLMFTNDYNLAQDYLASSOPTIALEMRDLOHMD 800
DB 834 EYSEMEDDTNGNIMEVEDS-----SKNLF-NDY-----PDSSVSANPEGESQ 876
DB 801 SALQIAKLHAPDQIPIFSKEVAIOLEFAGDYVNALAHVEKGITGDNKEHDEACL-----AG 856
DB 877 STIGAINPLAEE--YLSKODLLFL--DMLKFL-----CLCVTTAQ 912
DB 857 VAQMSIRMGDITRG-----VNOALKHPSR-----YLRDCCA-----ILENMQ 895
DB 913 TWTVSFRADIRKRLKMLLIDSSSTLEPTKSLHLMYLMKLKELPGEERYPLMEDVLELLKP 972
DB 896 FEEAQLYKGLYDKAASVYIRSKNMARVGDLLPHV-----SSPRIHLQYAKAKADGR 950
DB 973 LSNVCSLYR-----RDQDYCT--ILNHVILHVKNLQSNMDSNTRDAQOQ 1017
DB 951 YKEAVVAYENAKOWOSVIRIYDLHNNPRKAVNIYRETQSLDGAKMVAFFLOLDYGSA 1010
DB 1018 FLTVIGAF-----W-----HLTKERKYIFSVAM-----ALVNCLETLLEADPYSKW 1058
DB 1011 IQFLVMSK--CNNEAFT--LAQHNKMEIYADIIGSEDTTNDYOSIALYFEGEK-----R 1062
DB 1059 ALLNWMGKDFVNEVFTQFLADNHQVRLA-----ASINRLFQDFTGDDSR 1106
DB 1063 YLOA-----SKFLLCGOYSRALKHLKPCSSEDN-----VAIEALET 1101
DB 1107 LKALCPKIQOTAFENAYLAKAQBGRKSHSANEPTIDEIYRKSVALTLAVLSCSP 1166
DB 1102 VGQ-----ANDELTNQLIDLHIGE-----NDGMPKDAYIFRLYMAIK- 1140
DB 1167 ICEKQALFALCKSVKENGLEPHLVKRYLEKVSFTGYRLEDFPMASHLDYLVLEWLNQD 1226
DB 1141 -QYREAOFAIILIAEEDSAGNYRNAHD-----VLFSMYAELEKOKIKIPSE 1186
DB 1227 TEYNLSSEPFIL-----NYTNIEDFYRSCYVALLPHLVIRSHDFDKSIANOIOED 1278
DB 1187 MATNMLILSHYLVRX-----HYKGDHMKGARML----- 1216
DB 1279 KKSLLTDCPKRLVNLIPYFAVEGRDSGMAQOREATFVYMLKSENLLGQIDHLFIS 1338
DB 1217 -----IRVANNISKFPSHIYPILTSTYIECHR 1243
DB 1339 NLPETIVELLMTLHPANSSASQSTDLCDPFGSLDPAHPAPHPSPSHVYKATFAYISNCHK 1398
DB 1244 AGLKN-----SAFSFAMLMR--PEYRSKIDAKRK--KIE----- 1275
DB 1399 TKLKSLEILSSPSYQKILALICEQAETNNVYKKNHILKITYLHPUSLLKDKISGLG 1458
DB 1276 ---GWVRPDISIEEATTPCFC-----KFLPECELLCPGCKNSIPYCIATGRHML 1325
DB 1459 GAMAFVLRDVIYTLHYINQRPSCIMDYSLRFSJLC--COLLSQVQGTAVTYC----- 1509
DB 1326 KQDMTVCPHCDPFPALYSELKIMLNTESTCPMCSERLNAQAOLKIDCQOYL 1376
DB 1510 KD-----ALENHLHYIVGT--LIPLYEDEVQV--KQVLDLLKYL 1545

RESULT 10
US-08-642-274D-3
; Sequence 3, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO

```

; TITLE OF INVENTION: SCREEN FOR A PARTIAL A-T PHENOTYPE
 ; FILE REFERENCE: 22900033
 ; CURRENT APPLICATION NUMBER: US/08/642,274D
 ; CURRENT FILING DATE: 1996-05-03
 ; NUMBER OF SEQ ID NOS: 220
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 3056
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 ; US-08-642-274D-3

Query Match 1.8%; Score 131.5; DB 4; Length 3056;
 Best Local Similarity 17.6%; Pred. No. 0.098;
 Matches 221; Conservative 18; Mismatches 425; Indels 477; Gaps 60;

QY 323 VSQTLNKVATCGDNCIRIQ-----DLVDLKMVYIINL-----DEENKGLGTLMSWTDG 371
 DB 455 VLRCLEVALCQDKRSNLESSQSKDLKLNKIMCITFRGISSEKQAENFGLLGAIIOG 514
 QY 372 QLLALSTQKSLHVFELTKLPILGDACSTRAYLTSLELVANPVGCELPTIVSVDEPN 431
 DB 515 SLVEVDREFWKLTGSAACRSPAVCCLTALTTST-----VGAVMGMEION 562
 QY 432 FVAVGLYHLAVGNMNAFVIGENAVKRL-----KDMELGTVASICLHSDYAAALFE 485
 DB 563 MCEV-----NRS--FSLKESTMKWLLFQLEGDLNENSTEVPTI-LHSNFPILVLE 609
 QY 486 G-----KVQHLIES--EILDAQEEERETLFPAYD-----DKCRILCHALTS 525
 DB 610 KILVSLTMKCKAAMNPFQCEPHHKDEELSEVEFLIQTTFDK-----M 660
 QY 526 DELIYGTDRGVQY-----FYIE-----DMQFVNDYRHPVVKIIFPDNG 566
 DB 661 DELTIVRECEIEKHQSSIGSVHONIKESLDRCLLGLEQLNNYSEIT----- 710
 QY 567 TRLVFIDEKSDGFVYCPVNDATYEIPDFSPITKIVLMMNPMQGYEIVADDKVTYV 626
 DB 711 -----NSETLVRC-----SKLLVGL-----GCY-----CIVGVI 735
 QY 627 HKDTQCAKVIAGSTKVPFAHRLPYLNGELTCQTSQSKVNNIYLSHGLSNLMDXGP 666
 DB 736 AEEEAAYKSELFQKANSIMQACGESITLFKNTNEFFRIGSLRMNMOCLCTGSLNCTKSP 795
 QY 687 DELRP-MLANINMLKRFSDMCMCRILNDEAANNELARACLHMEVEFAIRVRRIQNG 745
 DB 796 NKIASGFFLRLLTSKLMNDIADICKSL-----ASFIRKPPDR---G 833
 QY 746 IVNSLE-----QIKGIEDVYLLAGHLAMFTNDYLNADLYLASSCPALAEEMRRDQHMD 800
 DB 834 EVSMEDDTGNLMEVEDOS-----SNKLF-NDY-----PDSSVSDANPESEQ 876
 QY 801 SALQALKHAPDOIPIFSKEAYIQLEFAGDYVVALAHYEGITGDKENDEACL---AG 856
 DB 877 STGAINPLAEE--YLSKODLLFL--DMLKFL-----CLCVTTAQ 912
 QY 857 VAQMSIRMGIRRG-----VQAALKHPSR-----VLRDQGA-----ILEMKQ 895
 DB 913 TMTVSFRAADIRKRLMLIDSSITLPEPKSLHLMYLMKLKELGEEVPLRMEVLELKP 972
 QY 896 FSEAAOLYEGKLYDYKAAVYIRSKNMAKYGDLPLHY-----SSPKIHLQYAKAKEADGR 950
 DB 973 LSNVCSLIR-----RDQVCKT--ILNHVLAHYVKNIGGOSNMSENRDQOG 1017
 QY 951 YKEAVVAAYENAKQMOVYIRIYLDHLNPERAVNIVRETQSLGAKVYARFQLDGYGA 1010
 DB 1018 FLTVIGAF-----W-----HLTKERYIFSVM-----ALVNLCKLTLEADPYSKW 1058
 QY 1011 IQFLVMSK--CNNEAFI--LAQOHKKMEIYADIIGSEDITNEDYQSLALFEBEGK-----R 1062
 DB 1059 AILNVMGKDPVNEVFQFLADNHQVRLA-----AESINRLFQDTKGDSR 1106

QY 1063 YLOA-----GKFFLGGQYSRALKHFLKCPSESDN-----VAIEMALET 1101
 DB 1107 LKALPLKIQQTAFENAYVLAQOGMRMSHSANPETLDEIYRKRSVLLTILVAVLSCSP 1166
 QY 1102 VGO-----AKDELLTNQILIDHLGE-----NDGMPKADYLFRLYALK- 1140
 DB 1167 ICEKQALFALCKSVKENGLEPHLVKKVLEKVSFTFGYRLDEDMASHLDVLEMLNLQD 1226
 QY 1141 -QYREAQRAIITAREQSGNRRNMD-----VLFMYAEKSKIKITPSE 1186
 DB 1227 TEYNLSFPFIL-----NYTNIEDFYRSCYKVLIPHLVIRSHFDEKXSIAQIOED 1278
 QY 1187 MATNMLHSHYILVKI-----HVKGDHMKGARML----- 1216
 DB 1279 WKSLLDPCFKLIVNLPFATYEGTSDSGMAQORETATKYMDKSENLLGKIDHFLFS 1338
 QY 1217 -----IRVANNISKEPSSHIVPLTSTVTECHR 1243
 DB 1339 NLPEIVELMLTLEPANSASQSTDLCDPSGDLDPAPNPHPSPHVIKATFAYISNCHK 1398
 QY 1244 AGAKN-----SAPFPAALMR-PEYRSKIDAKYK--KIE----- 1275
 DB 1399 TKLSILEILSKSPDSYOKITLAIQCAAEETNNVYKRRILKIYHLFVSLLDKIDKSGLG 1458
 QY 1276 ---GMYRRPDISEIEEATFPCCPC-----KFLPECELLCPGCKNSIPCIATGRHML 1325
 DB 1459 GAAAFULRDVITYLTIHINRPSGIMDVLSRSLC-CDLLSYOCQATAYTC----- 1509
 QY 1326 KDDWTVCPHCDPPALYSELKIMLNTSTCPMCSERLNAOQLKISDCTOYL 1376
 DB 1510 KD-----ALENHHLIVIGT--LIPLYVEQVEVQ--KQVLDLKLVL 1545

RESULT 11

US-08-952-127-3
 ; Sequence 3, Application US/08952127
 ; Patent No. 6211336
 ; GENERAL INFORMATION:
 ; APPLICANT: Shiloh, Josef
 ; APPLICANT: Tagle, Danilo A.
 ; APPLICANT: Collins, Francis S.
 ; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Kohn & Associates
 ; STREET: 30500 No. 6211336thwestern Hwy., Suite 410
 ; CITY: Farmington Hills
 ; STATE: Michigan
 ; COUNTRY: U. S.
 ; ZIP: 48334
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/952,127
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kohn, Kenneth I.
 ; REGISTRATION NUMBER: 30,995
 ; REFERENCE/DOCKET NUMBER: 2290,00029
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 810-539-5050
 ; TELEFAX: 810-539-5055
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3056 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein


```

OY 118 CIYLDAMTNRKTSQLDNGMRDQ---MSFLLWS-----KVGS--FLAVGTVKGNLXIYHQ 167
| | | | | : : : : : | | | | | : : : : :
DB 209 TIRLMDPKSGQC--LGDALRGHSKWITSLSWEPILHLVPGSKPRLASSSKDGTIKIMDTV 266
| | | | | : : : : : | | | | | : : : : :
OY 168 TSKRIPIVLGKHTKRTTCGCGMAENLXALGEDKMTIV--SNOEGDTIRQTVR----- 218
| | | | | : : : : : | | | | | : : : : :
DB 267 SRVCCYTMSGHTNSVCYKMGOGLLYSGSHDRTVRYVMDINSQRCINILKSHAHMVNHL 326
| | | | | : : : : : | | | | | : : : : :
OY 219 -----SEPXNMQPFLLM-----KMDRTSAESMISVGLKKTLEFL 254
| | | | | : : : : : | | | | | : : : : :
DB 327 SLSTDYALRIGAFDHTGKRPSTPEBAOKKALENYEKICKNGNSEEMWTASDDYTMFLW 386
| | | | | : : : : : | | | | | : : : : :
OY 255 NINEPDNP-ADLEFQODFGNIVCYMWYGDGRIMIGFSGHFVYISTHTGELGOEIFQARN 313
| | | | | : : : : : | | | | | : : : : :
DB 387 NLKSTKPIARMTGHOKLVNHAFFS--PDGRYIVASAFDN--SIKLMGDRGCKFISTFRG 442
| | | | | : : : : : | | | | | : : : : :
OY 314 HKDNLSTIAVSQTLNKVATCG--DNCIKIODL-----VDLKMVIYLLNDEENKGLGTLS 366
| | | | | : : : : : | | | | | : : : : :
DB 443 HIASVYQVAMSSDCRLVSCSKDTTLKWDVTRKRLSVDLPIKTKLYVD----- 492
| | | | | : : : : : | | | | | : : : : :
OY 367 WTDGQ 372
| | | | | : : : : : | | | | | : : : : :
DB 493 WSVDC 498
| | | | | : : : : : | | | | | : : : : :

```

```

RESULT 14
US-08-477-346-66
: Sequence 66, Application US/08477346
: Patent No. 6262023
:
: GENERAL INFORMATION:
: APPLICANT: Mochly-Rosen, Daria
: TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
: NUMBER OF SEQUENCES: 265
: CORRESPONDENCE ADDRESS:
: ADDRESS: Morrison & Foerster
: STREET: 2000 Pennsylvania Avenue, NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20006-1812
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/477,346
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/487,072
: FILING DATE: 07-JUN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: MURASHIGE, KATE H.
: REGISTRATION NUMBER: 29,959
: REFERENCE/DOCKET NUMBER: 2550-0025.20
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1500
: TELEFAX: (202) 887-0763
:
: INFORMATION FOR SEQ ID NO: 66:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 514 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: INDIVIDUAL ISOLATE: YCW2 PROTEIN, Fig. 49
:
: US-08-477-346-66

```

```

Query Match 1.8%; Score 131; DB 4; Length 514;
Best Local Similarity 20.9%; Pred. No. 0.0043;
Matches 89; Conservative 57; Mismatches 164; Indels 116; Gaps 20:

OY 4 IGRSGRLGIDGNAVACLVETGCACVLHSSRSGSVEMKRIFSLLEKTLGAPIOPAMOK 63
| | | | | : : : : : | | | | | : : : : :
DB 132 YTRSSAIIAGHSITLIC-----SAFAPHTSSR----- 158
| | | | | : : : : : | | | | | : : : : :
OY 64 TSGNLAIVTGA-DYIVKIFDRHGQKRSEINLPGN-----CVAMDMDKGDVLAVIAEKSS 117
| | | | | : : : : : | | | | | : : : : :
DB 159 -----MYTGADNTRARLMDCDTQTPMH-TLKGHYNWLVS--WSPGEVIAT-GSDMN 208
| | | | | : : : : : | | | | | : : : : :
OY 118 CIYLDAMTNRKTSQLDNGMRDQ---MSFLLWS-----KVGS--FLAVGTVKGNLXIYHQ 167
| | | | | : : : : : | | | | | : : : : :
DB 209 TIRLMDPKSGQC--LGDALRGHSKWITSLSWEPILHLVPGSKPRLASSSKDGTIKIMDTV 266
| | | | | : : : : : | | | | | : : : : :
OY 168 TSKRIPIVLGKHTKRTTCGCGMAENLXALGEDKMTIV--SNOEGDTIRQTVR----- 218
| | | | | : : : : : | | | | | : : : : :
DB 267 SRVCCYTMSGHTNSVCYKMGOGLLYSGSHDRTVRYVMDINSQRCINILKSHAHMVNHL 326
| | | | | : : : : : | | | | | : : : : :
OY 219 -----SEPXNMQPFLLM-----KMDRTSAESMISVGLKKTLEFL 254
| | | | | : : : : : | | | | | : : : : :
DB 327 SLSTDYALRIGAFDHTGKRPSTPEBAOKKALENYEKICKNGNSEEMWTASDDYTMFLW 386
| | | | | : : : : : | | | | | : : : : :
OY 255 NINEPDNP-ADLEFQODFGNIVCYMWYGDGRIMIGFSGHFVYISTHTGELGOEIFQARN 313
| | | | | : : : : : | | | | | : : : : :
DB 387 NLKSTKPIARMTGHOKLVNHAFFS--PDGRYIVASAFDN--SIKLMGDRGCKFISTFRG 442
| | | | | : : : : : | | | | | : : : : :
OY 314 HKDNLSTIAVSQTLNKVATCG--DNCIKIODL-----VDLKMVIYLLNDEENKGLGTLS 366
| | | | | : : : : : | | | | | : : : : :
DB 443 HIASVYQVAMSSDCRLVSCSKDTTLKWDVTRKRLSVDLPIKTKLYVD----- 492
| | | | | : : : : : | | | | | : : : : :
OY 367 WTDGQ 372
| | | | | : : : : : | | | | | : : : : :
DB 493 WSVDC 498
| | | | | : : : : : | | | | | : : : : :

```

```

RESULT 15
US-08-473-089-66
: Sequence 66, Application US/08473089
: Patent No. 6342368
:
: GENERAL INFORMATION:
: APPLICANT: Mochly-Rosen, Daria
: TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
: NUMBER OF SEQUENCES: 265
: CORRESPONDENCE ADDRESS:
: ADDRESS: Morrison & Foerster
: STREET: 2000 Pennsylvania Avenue, NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20006-1812
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/473,089
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MURASHIGE, KATE H.
: REGISTRATION NUMBER: 29,959
: REFERENCE/DOCKET NUMBER: 2550-0025.22
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1500
: TELEFAX: (202) 887-0763
:
: INFORMATION FOR SEQ ID NO: 66:

```


SEQUENCE CHARACTERISTICS:
LENGTH: 514 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: YCW2 PROTEIN, Fig. 49
US-08-473-089-66

Query Match 1.8%; Score 131; DB 4; Length 514;
Best Local Similarity 20.9%; Pred. No. 0.0043;
Matches 89; Conservative 57; Mismatches 164; Indels 116; Gaps 20;

QY 4 IGRCSRLGLDGNVAVCLVETGACVYLHSSRSSGVEMKRIFSLLEKTLGADIPQFAMOK 63
DB 132 VTRSSSNAIGHGSTITL-----SAFAPHTSSR----- 158
QY 64 TSGNYLAVTGA-DYIVKIFDRHGOKRSEINLPEN-----CYAMDMDKGDVLAIVAEKSS 117
DB 159 -----WVTGAGDNTARIMDCDTOTPMH-TLKGYMMVLCVS--WSPDGEVIAT-GMDN 208
QY 118 CIYMDANTNKTSQLDNGKRDQ---MSFLMS-----KVGS--FLAVGVKGNLXIYNHQ 167
DB 209 TIRLMPKSGQC--LGDALRGHSKMTSLSWEDIHLYKPGSKPRLASSSKDGTIKIMDTV 266
QY 168 TSKRIPIVLGKHKRTKTCGGCNAENLXALGGEKMTIV--SNOEGDTIRQTQVR----- 218
DB 267 SRVCQYTMGHTNSVSCVKGCGLLYSGSHDRTVWMDINSQRCINILKSHAHVYVNL 326
QY 219 -----SEPXNMQFLM-----KMDRTSAESGISVVLGKKTLEFL 254
DB 327 SLSTDYALRIGAFDHTGKKPSTPEAOKKALENYEKICKKNGNSEEMVITASDDYTFLW 386
QY 255 NLNEPDNP-ADLEFODDFGNIVCYNMYGDRIMIGFSCGHFVVISTHTGELGOEIFQARN 313
DB 387 NPLKSTKPIARMTHQKLVNHVAFS--PDGRYIVSASFON--SIKLMGDRGKFISTFRG 442
QY 314 HKDNLTSIAVSQTLNKVATCG-DNCIKIODL-----VDLKDMYVILNLDEENKGLGTL 366
DB 443 HIASVYOVAMSSDCRLVSCSKDITLKVMDVRYTRKLSVDLPGIRKTLKYD----- 492
QY 367 WTDDGQ 372
DB 493 WSYDGK 498

Search completed: May 21, 2002, 14:50:30
Job time: 50 sec

THIS PAGE BLANK (USPTO)

Db 361 NAINILFK--NGDENCVALAVDCGKSKDKTLNKLKVELLGE-DGNVKDPAQLEFLRV 417
Qy 1138 ALKOYREAAQTAAIITAREEOAGNYRNADVLFSMYAEELKSOKIKIPSEMATNLMIHSY 1197
Db 418 GIGRRKDAQTAQVAVVAQIHQAGKNYRIARDLFLQHQOLREKMRPIPLDMKNSLMAIHSY 477
Qy 1198 ILVKIHKVNGDHMKGARMLIRVANNISKFPS 1228
Db 478 IIVKALINKEFTLLAARLLIRTCGEIQRPT 508

RESULT 2
T27881
hypothetical protein ZK520.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27881
R:Steward, C.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z20434
A:Accession: T27881
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1 634 <MIL>
A:Cross-references: EMBL:Z29822; PIDN:CAB07301.1; GSPDB:GN00021; CESP:ZK520.3
A:Experimental source: clone ZK520
C:Genetics:
A:Gene: CESP:ZK520.3
A:Map position: 3
A:introns: 53/2; 75/2; 120/2; 187/1; 318/2; 341/1; 455/1; 530/3

Query Match 12.7%; Score 924; DB 2; Length 634;
Best Local Similarity 30.4%; Pred. No. 3.8e-51;
Matches 204; Conservative 130; Mismatches 239; Indels 98; Gaps 8;

Qy 56 PLOFMAQRTSGYLAVTGADYIVKIFDRHGQKSEINLPGNCVAMDMDKGDVLAIVAEK 115
Db 37 PLIHRM-RENGHTVAVACANNNTVIYDKKGNIDALNPTGKLIDIAMDEGDVLAIVAN 95
Qy 116 SCSTILMDNNTKTSQOLDNG--MRDMSFLMSKVSFLAVGTGKGLXIIYNHOTSRI 172
Db 96 TCTITLMDVNSKNTDVEGATSSKEPLTCLAMSPSTPLVIGNNAAGNIYVNNHRTSRI 155
Qy 173 PYLGRHTRITGCGWNAENLXALGGEKMITVSNQEGDITROTQVRSPEXNMQFLMKMD 232
Db 156 AVMGKHQRSVTOITVTPEDYVISCSDNTLSVTLEGTVSTTTNGEPTNDY--GSVN 213
Qy 223 DRTSAESMISVVLGKTFELNLNPNADLEFOODFGNIVCNWYDGGIMIGFSCG 292
Db 214 GKGSGGVTVMSVYIGKTKILMAHYNALDEPNMLQDEKGNHSHYRMFNDGYILIGFDRG 273
Qy 293 HFNVTSTHTGELGQEFQARNKDNLTSAVSGQTLNKVATCGDNCIKIDVLDMYVI 352
Db 274 YIISISAHNNEGSELYSLERGLASTAVSTSFNKLITIDNMVKKVADDELITVYML 333
Qy 333 INLDEENKGLTSLWTDGQLLALSTQGRSLHVFLLKPLIDACSTRIVLTSLEVTY 412
Db 334 TEIETE-KNLSEIETVEDQOLVAVSSQSGVLSIFVTKMPTLAAYNSICYLTNLTOVTY 392
Qy 413 ANPVGELPIYVSVDPERFVAVGLYHLAVGNMNRMF-----VLGENAV 458
Db 393 VAEVREKGSSTLELNIETPLMGLPNI LAVANNNTVFEDYHTPAOMQAQQLQSTQSA 452
Qy 459 KK-----LKMEYLGTVASICLHSDYAALFEKGVOHLIESEILDAOERETRLFP 510
Db 453 EKPTLYAAEPINREYLSVTYNIQLNYMAAVNFGSRLHLIRN-----SDNVSIERP 507
Qy 511 AVDDCRILCHALTSDFLIYGTDTGVVOYFIEDMQFVNDYRHPVSVKKIFPDPNGTRLV 570
Db 508 EANNRATLYSVALTENEL----- 526
Qy 571 FIDKSDGVYCPVNDATYEIPEPSFTIKGVLMENMPMDKGVFIAYDDKVTYTVFHKDT 630

Db 527 -----FTTSYKGAIMEFTTIDKNFAVFDSONIVFLLSKQH 563
Qy 631 IOGAKVILAGSTKVPFAHKPLLNYNGELTQCQOSKVVNNIYVSTFGPLSNLKDXPDELR 690
Db 564 IOGESVITYSATRLPHAVPLSLNKGIVTCLMSNGLSLSDSHKTESVLSDKSEYID 623
Qy 691 PMLAHNMLKR 701
Db 624 DILTRSLLMHR 634

RESULT 3
T34393
hypothetical protein T27B1.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34393
R:Wu, X.; Le, T.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid T27B1.
A:Reference number: Z21517
A:Accession: T34393
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1758 <WUX>
A:Cross-references: EMBL:U41020; PIDN:AAA82332.1; CESP:T27B1.1
C:Genetics:
A:Gene: CESP:T27B1.1
A:introns: 13/3; 61/3; 99/2; 134/3; 161/2; 186/3; 251/3; 391/3; 571/3; 650/3; 747/3;

Query Match 3.7%; Score 270.5; DB 2; Length 1758;
Best Local Similarity 18.7%; Pred. No. 1.3e-08;
Matches 270; Conservative 220; Mismatches 493; Indels 463; Gaps 68;

Qy 65 SGNVLAIVGADYIVKIFDRHGQKSEINLPGNCVAMDMDKGDVLAIVAEKSCITYLMDA 124
Db 27 NGSRAAIACSDRSVALDLBENGVOKDRF---TCKPID-----AKYGKSKFTVLC--- 71
Qy 125 NNKNTSQLDNGMRDMSFLMSKVSFLAVGTGKGLXIIYNHOTS--RKIPYLGKHTK-- 180
Db 72 -----MTFSPDSSRIALIGOSDNVLFPIYKGTWMEKRYIVNKFVOPS 113
Qy 181 RITCGCWNAENLXALGGEKMITVSNQEGDITROTQVRSPEXNMQFLMKMDRTSAES 240
Db 114 AVTCLSWPDDKILVGOQDQKVRIG---LTKTKCS-----LYKTD-----ET 154
Qy 241 MISVVLGKTFELNLNPNADLEFOODFGNIVCYNMWYDGGIMIGFSC----- 291
Db 155 VVSITQTHPRKRTSFVSAH-----OD-GSIILYNFSSTQSKI---CTLOVPYNYL 199
Qy 292 ---GHFVYS-----THTGELGQEFQARNKDNLTSAVSGQTLNKV--ATCG--- 334
Db 200 VFTNGLIVATSDRVLSTENGVOQDFYNDQSEKESISSCPTAONVYGTGCGGVF 259
Qy 335 --DNCK-----IDLVND-----LKDMYV-- 351
Db 260 SVDCLLRGMLKSRPETYVAPSHVILRDVTNDRTVNSNGLAIDELKIMGXRYIYG 319
Qy 352 -----INLDEENKGLTSLWTDGQLLALSTQGRSLHVFLLKPLIDACSTRIVLTS 406
Db 320 YTSSSIILADSEQSFSELEW-----QSGHKEFFD--FNCC-----LIIN 360
Qy 407 LLEVIVAN-PVGEPLPIYVSVDPERFVAVGL---YHLAVGNMNRMPYVLGENAVKKLK 462
Db 361 AGEVTVVEYGVDSIGWRTSLTSPHLSVQKKAFFSYGNFNS-----ESSVSYSE 414
Qy 463 DMEYLGTVASICLHSDYAALFEKGVOHLIESEILDAOERETRLFPAYDCKRI---- 518
Db 415 HSDL--NGQICVFSDFL-----QRTSSVFFLIKAKKIFNDL 451
Qy 519 -LCHALTSDFLIYGTDTGVVOYFIEDMQFVNDYRHPVSVKKIFPDPNGTRLVFIKSD 577

```

Db 452 QLEHVNIPVLLINGQOESFINHTGAVDMIELNE-----RASKLLYRDKRS- 496
QY 578 GFVYCVNATYFRIPEPSPIKGVL-----WEMPMKGVFIAYDDKVTYTFPHKD--- 629
Db 497 -----KVLTVIISDQSRVLLSFCTYQVWPMSDVYVQSGDNLSIW-YNDPLPE 545
QY 630 -----TIQC-AKYLIGSTFVPPAHKPLLLYNGELTCQOTOSGKV-----NNIYSTHGFLS 679
Db 546 QVTNMIKIGVEAVLADADRT-----EVIQVETAVAYVELDNTQIE---FGA 590
QY 680 NLDKXDPDELRPMLAHNLMKRFSDAMEMCRILNDEAAMNELARACLIHMEVEFAIRVYR 739
Db 591 ALEKRFEDRAVAFLESN---TSGDAYSM-----WIRVEMALEHGNLFVAQCVCYA 638
QY 740 RIGNVQIVWSLEQIKGED-----YVLLAGHLAMFNQDNLADLYLASSCPI 787
Db 639 AINDVAKVRRLHDLITADEASISIGDGTHTFKVRAMLAIMGRKFEARIFLEQNDTE 698
QY 788 AALEMRDLOHMSALQALAKHLAPDQIPFISKEY-----AIOLEFA-GDYVNA 834
Db 699 SAIGMTSLHKDEALELAKVLMPEYEQLTSLALSTGDSKRAELKVSDDGTLSA 758
QY 835 LAHYEK-----GITGDNKE---HDEACLAGVAOMSIRMGDIRGVNQLKHPSRVLR 884
Db 759 IOLYKSNKPLSALSAANDSVLSQDENILRQIADSLVK-----SOLYD 802
QY 885 DCGAILIENKQSEAAQLYEKGLYDKA-----ASYIIRSKMNAKVGDLPHVSSP 935
Db 803 KADVYEKLDKDFKAVEYFKKGDAYGKAIQIARAFPEKVVTLLEQENG----- 850
QY 936 KIHLOFAKAKEADGRY-----KEAVVAYENAKOMOSVIRITLDHLNKEKAVNIYRET 988
Db 851 -LLEITIGQYDAVNVHFEVNDLKKAVEAIIKKEW-----PKALSTV--- 892
QY 989 QSLDGAKM-----VAREFLQADYGSALQIOLFLMSKNNEAFTLAQOHNK---MEIYA 1037
Db 893 ENIQDQKVRGTGYGEIADHYSNKGDPERABRLFEVAGLFNDALIMYGKNNKWIADFLTSE 952
QY 1038 DIIGSEDTTNEQDSTALTYEGEKRYIQAGKFFLLCGOYSRAKLNKPLKCSSEDNVAIEM 1097
Db 953 EFGGRATIS-SYLAKAEDLDHGRFAEAEOLYITIGMPKAIOMY----- 997
QY 1098 AITFVQAKDELITNOLIDLGENDMPRDAKYLFRLYMAKQVREAAQTATIIAREEQ 1157
Db 998 --DRVGRDDVY---RLVEYRGEH--MHETRRKF-----ATQYERGD---LKAABEQ 1041
QY 1158 --SAGNYRNAHDVLFSSY--AELKSOKIKI-PSEMATNL--MILHSY-----ILVKIH 1203
Db 1042 FLKAGDFRSA---VNMKYSEMMSDAYRIAKTEGGENMEKQVLFEMMAKSIGDDAAVKTL 1097
QY 1204 VKRGDHMKG-----ARMILRVANNISKFPESHYPIILSTYIECHNRAGLKNSAF 1251
Db 1098 NKGMLEMGIDFACETGAFLADFLARIG--AKDRMGTAVHRLATQLE--EGRELEASK 1153
QY 1252 SFPAMLMRPEY-----RSKIDAKYKKIKEGMVRPRDISEIEEATTPCPC 1296
Db 1154 HYVEAIKLNNTYTTCQAVPSRFDLQKRLGNKPELAVEHFIINDNWADAERYAK--DHC 1211
QY 1297 KFLLEPE 1302
Db 1212 ESTLPPD 1217

```

RESULT 4

T43484
 hypothetical protein DKFZp434K016.1 - human (fragment)

C:Species: Homo sapiens (man)

C:date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000

C:accession: T43484

R:coauthor: K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, December 1999

A:reference number: 22514

```

A:Accession: T43484
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1198 <AAA>
A:Cross-references: EMBL:AL133565
A:Experimental source: adult testis; clone DKFZp434K016
C:Genetics:
A>Note: DKFZp434K016.1

Query Match      3.6%; Score 263; DB 2; Length 1198;
Best Local Similarity 18.2%; Pred. No. 2.2e-08;
Matches 245; Conservative 197; Mismatches 506; Indels 396; Gaps 54;

QY 116 SSCITVLDANTNKTSQLDNKMRDMSFLMSKVSFLAVTGVGNLXITHYQSRKIPV- 174
Db 70 SSDFGLMSPQKSVSK--HKSSKIIICCSVFNQGYALALMENGILSTIRKNKEBEVKIE 127
QY 175 -LGKHTKRTTCGCM-----NAENLXALGEDKMI-----TVSNOEG 209
Db 128 RFGSLSPIWISICMNPSSRRESFMMNRENDA---EDVIYNRVYIQEIPSLKSAVYSSOG 184
QY 210 DTRQTVQVSEPYNMQFLLKMKMDRTSAESMISVY-LGKKTFF-LNLNEPNDLDEF 267
Db 185 SEAEERPEEEDSDSPR-----DNLLEERNDDILAVADMCKVSEFYQLSGKQICKDRALNF 238
QY 268 QDPGNTVCYNNYGDR--IMIGFSCGHFYVISTHTGELGQEIFQARNHKNLTSIANSQT 326
Db 239 DP-----CCISYFTKGEYILG---GSDKQVSLFT-----KGV----- 269
QY 327 LNKVATCGDNCIKIQLDVLDKMAYVILNDEENKGLTSLSMDDGOLLALSTQSLHVF 386
Db 270 --RLGTGV-----EQMSWWTQCAKSDSNVYVVGCGDDGTISFY 305
QY 387 LTKLPLIGDACSRIAYLISLEEVYANPYEGELPTVSVDPENFVAGVLYHLAVGMNN 446
Db 306 QLIFSTVHGILKDRYAVRDSMTDVIYQHLLT--EQKVRIRKCELVKIAIYRNRLADLP 364
QY 447 RANFYVLGENAVKIKLDMETLGTVASICLSQVAAALFEKQVQHLIESIILDAOEER-E 505
Db 365 KILYELYS--EDLSDMH-----RVKEKTIKKFECNL-LVVCANHIIICOEKRLO 412
QY 506 TRLPFAVDK-----CRILCHALTSDFLIYGTGTGVQYFLEDMQFVNDYRHPV 555
Db 413 CLFSFGYKEREMQMESLIRIKYIGPRGREGILLVGLKNGQILIKIPVDNLFATVLLKQAT 472
QY 556 SVKKIFPDNGRLVFIIDEKSDGFVYCPVNDATYEIIPDSPTIKGLVEMPMKGVFT- 614
Db 473 AVRCIDMSASGRKKLAVVDENDTCLVY---DIDRKELFQEPNANSVAMNQCEDMCLFSG 529
QY 615 -AYDDQKVTYTVPHKDTIQCAKYILAGSTKVPFAHKLILLYNG-ELTQCOTOSGKVNIIYL 672
Db 530 GGYLNTIKASTFVPHROKLOQ--FVVG-----YNSKILFC----- 561
QY 673 STHGFLSNLKDQXDPDELRPMLAHNLMKRFSDAMEMCRILNDEAAMNELARACLIHMEVE 732
Db 562 -LHVF--SISAVEPQAPRYQY-LDRKLFKFAVQIACLVTDIMELAMELEGLDFE 617
QY 733 FAIRVYRRIGN---VGIVMSLEQIK--GIEDYNLLAGHLAMFNQDNLADLYLASSCPI 787
Db 618 TARKAFIRYQDLYRLLELSSIEERKKGTNNDLFLADVSYOGKFHEAKLYRSGHEN 677
QY 788 AALEMRDLOHMSALQALAKHLAPDQIPFISKEVAILOEFAGDYVNNLAHYEKGTGDNK 847
Db 678 LALEMTYDL-----CMETAKDPLG-----SGDPK 702
QY 848 EHDACLAGVAOMSIRMGDIRGVNQLKHPSRVLRKDCGAILIENKQSEAAQLYEKG 907
Db 703 E-----TKMLITQKADAR-----NIKEPAAYEMVYISAG 732
QY 908 YVDAKASVYIRSKNAKVGDLPHVSSPKTHLOFAKKEADGRKKEAVVAYENAKQOSV 967
Db 733 EHVKAIEI-CGDHGWV---DMLIDI-----ARKLDKAREPFL-----L 767

```

[illegible]

Query Match	3.5%	Score 256.5	DB 2	Length 1462
Best Local Similarity	17.3%	Pred. No. 7.8e-08		
Matches 247	Conservative 219	Mismatches 485	Indels 475	Gaps 64
OY	54	GAPIDFAWOKRTGNYLAV-----TGADYIYKIFDRHGO--KRSEINPGNCVAMDMDK	104	
	: :	: :	: :	
Db	16	GSPFSLSNMHPH-PLPLAVAYISTTSGS---VDIYLEGECYCPDITVERPFRVASLCHMP	71	
OY	105	DGDVLAVIAEKSSCIYLDAMNT---NKTSQLDNGH---RDMOSFLMKSGVSLFVGT	156	
		: :	: :	

Dd	72	TRVLAV-----	GMERGEVTVFNKKOKEDQHTMPLHTADITVLKMSPSGNCLLSD	122
Qy	157	VKNGLXINYHOTSRTK---	PVL-----GKHKRTITCGCW-----NAENL-----XALGED	199
Dd	123	RLGVLLRLRLRLOQRVVGST	PLIKHEXKHLNTH-----CIFRLPPGEDLVLAKKAAVSGDE	178
Qy	200	KMIIVSNOEGOTINOTQVRSP	RMQFELKMD-----DRTAASMTISVGLKKTLEF	253
Dd	179	KALDMFNKKSSSGSLKMGSH	EGLEFVSLSMOGVYHYDEKKTTOVVSAUSTIOMLEY	238
Qy	254	INLNB-----	PNPAD-----LEFOODEFNICYNMWYDGRIMIGFSCG	292
Dd	229	MEKKRALVYVTEENRLSLY	TPRPGKAEVMMYKLSKTCR-----RADIALLEG	288
Qy	293	HEVVISYHTG-----	ELGOETFOARNKH-----DNLSIAVSQTLNKVATCGDN--	336
Dd	289	SLLVMAVEALRPMWDE	RGENTYILSPDEKFGFEKGENMNCVCYCKVGLLAAGTDRGV	348
Qy	337	--CIKIDOLY-----	DLKDMVILINDEENKGGTLSMTDDGOLLASPOKRSIHYELTK	389
Dd	349	AMMKRVPEPLSPGAEGRD	RALOTPELOGNTIQIOMGSRKKLLAVNS-----VTS	400
Qy	390	LPILAD--ACSTRIALY	TSLEVTYANPEGELPITVSVDPENFVAVGLYH-----	439
Dd	401	VALLSERAMSSHFHOOVA	MQVS-----PSLNV-----CELSYGAHSLRTRDMHS	447
Qy	440	-----LAVGMNNRA	FYVLGENAVKIKLDMELYGTVASICLHSDYAAALPEGVOLH	491
Dd	448	GVFATKADVAVMNROVA	IEFELSGAALRSAG--TFICLETPLVLMHEENYTVESNNVOYR	505
Qy	492	LIESIILDAOEBERTRL	EPVANDOKCRILCHALNSDFLITGTDGVVQVYIIEDMQFVNDY	551
Dd	506	TWQGTVKOLLFSETEGN	PCFLDLC-----GNFLVGTDLALFSPDLRSRE--AKAH	556
Qy	552	RHPVSVKKIFP-----	DPKQTRLYEIDEXDSG-----FVYCPVNDATYEIPEIDSP	596
Dd	557	CSCRSIALBYGVGSIAS	LKSSSGSITSLPESKADNSPDKICFYDEMVDYVYDF--	614
Qy	597	TIKGVLMWMBMDKGV	FAYDDDKVYUYVYHNKDTIGAKAVIAGSTKVPFAN-----KPL	651
Dd	615	-----KTGIDRR	ETLSENEOE--TNKSHLEFVDEGLK-----NYVPLNHFMQSEPR	659
Qy	652	LLY-----	NEELICQOQSGK-----VANITYLS--THEFSLNKDXPD	687
Dd	660	LEVCEAVOETPRSQP	OSANG-----QPDQGRAGRAPADVILISFISEHGL--LHSEFP-	712
Qy	688	ELRPLAENL---	-----LKRF-----	702
Dd	713	--RRATSHSLGME	VPPYUFTKRKPELEADRDEVEPRCGCHNIPQVNSRRPLRDEVGLEDCK	770
Qy	703	-----	SDAMEKCRILNDEAAMNELARACLHMHMEVEFALIVYRIGNV	744
Dd	771	ATRDAMLHFSEFV	ITIGMDAEFAKSIKLKISEAVEMNNARMCVYQTRLDVA--KVC--LGNM	827
Qy	745	GIWVSLBDIKIETDY	NLLAGHLAMFTDYMLADDL--YLASSCPIALBEMKRLDQH--	798
Dd	828	GHARCARALREAEDEP	ELEARVALATOLCMLDABOLKRC-----KRHDLKKFYOA	881
Qy	799	---MDSALOKHLAP	QOIFISKEVAYIOLEFAGDVYNALAHKEGITGDNKENDEACLA	855
Dd	882	AGRMOEALOVAEHND	RHYLHSTYHRTAGHLEASADCSRALSTYKSDT-----	929
Qy	856	GVAQMSIEMGD	IRGVNAQALKHPSRVLYLKRPGCAILLENMKOFSEADOLYEKGLUYDKASV	915
Dd	930	-----	HREVP--RMLSEDLPSL-----ELVYVK-----	951
Qy	916	YIRSKN---	WAKVGLLLPHVSSPKIHLQYAKAEADGRYKEAVYAVENAKOMOSVIRY	971
Dd	952	--MKDKTLMRWMA-----	QYL-----ESQGEDAALHYELARDBHFSVLRH	991
Qy	972	LDHLNPEKAVNIYRE	QUSODGAKMARARFPLDGLDYSALORFLYMSKCNNEFATLQOHN	1033
Dd	992	C-FOGNVOKAOLAN	ETNGNLAASTYHALAROTSEOEVEGVAVHFYTRAQDAFKNIRLCKCN-	1049

A:Gene: CESP:C27A7.4
 A:Map position: 5
 A:Introns: 25/1; 53/3; 94/2; 158/3; 266/1; 354/3; 414/1; 523/3; 595/1; 631/2; 758/2; 103

Query Match 2.9%; Score 209.5; DB 2; Length 1307;
 Best Local Similarity 17.6%; Pred. No. 6; 7e-05;
 Matches 248; Conservative 222; Mismatches 500; Indels 439; Gaps 62;

```

OY 100 MWDMDGDLVAVIAEKKSGCIYLM-----DANTNKTSLDGMNDPOMSELMKSVKGSF- 151
    || || || || || || || || || || || || || || || || || || || || || ||
Db 39 VMSHSDGYVLMPLHNPSS-VHLXYAVWIGEELSTNIMQIE--LNDQIT--LMCKRLSTD 93
OY 152 ---LAVGTWKGKGLIYHQTSRKIPVLGKHTKRITCGCMAENLXALGEDKMITVSNO 207
    || || || || || || || || || || || || || || || || || || || || || ||
Db 94 KRPKPSISYTDSDGVDSPSGSKESLARRDEKSTVP-KGTEFLFA-GKSGTIGVND 151
OY 208 EBDTIRQOVRSRPNMQ-----FELMKMDRKSAAASMTSVVLGKKT 250
    || || || || || || || || || || || || || || || || || || || || || ||
Db 152 RORTIH--KLDESLIFMGYCEITSIITVPTROCFIHLAKGTSEMKCAERVVKLGKSE 209
OY 251 LEFLNLNEPDNPADLEFOODFGNIYCYMNGDIRIMIGSCGFVYISTHTELGCEIRFO 310
    || || || || || || || || || || || || || || || || || || || || || ||
Db 210 KSLSELD-----DGLVMC--YGEKEIRV-----WDLIKEENGTIALDVNK 248
OY 311 ARNHKNDLTSIAVS---QTLNRYATC-GDNCIKIODLVLDKMYVILNDEENKGLTSL 366
    || || || || || || || || || || || || || || || || || || || || || ||
Db 249 GFQPEETIIVWTVNGKRESLFFVYSLIRWSPILSTAALITEEDLVLI-----GENSLT 300
OY 367 WTDGOLLALSTQKSLAHFLKPLTLDGACSTRIAVLSLELV--VANPYEGELP--- 421
    || || || || || || || || || || || || || || || || || || || || || ||
Db 301 VKMRGMALITSSNSF-----TLHATISGVSDIKLSTIPSAK 338
OY 422 -----ITVSDVEPNFVAVGLYHLAVGNNRAMPVYLGENAARKLKDMEX 466
    || || || || || || || || || || || || || || || || || || || || || ||
Db 339 GICLGEKOLVYMNEDPVTYVYDQTSLATIQCTSFSCNTTSA--IVNOLKCIERKDF 395
OY 467 LGTVASICLHSDYAAALEFGKVOHLIESEIIDAOERETRLFPAYDDKCRILCHAITS 526
    || || || || || || || || || || || || || || || || || || || || || ||
Db 386 ARTL-----OGTLROEISLPEIERVVLIQIMOLKIFQOVS-----TVMER 437
OY 527 FLIYGDGCVQGYFI-----EDMOFVNDYRH-----PVSVKKIFPPNPTRLVLEFI 574
    || || || || || || || || || || || || || || || || || || || || || ||
Db 438 LLYVAELDAVSYSFSDKGMTDTQETVOELAHITSSGPVTAARKKARQDSRQOMNH 497
OY 575 KSDGFVYCPVNDATYEIPDFSPITGVLMENPMKGVFIAYDDKVYTYVEHFKITQGA 634
    || || || || || || || || || || || || || || || || || || || || || ||
Db 498 RRGAEEM--DENARLYL-----VECIHVEPESTDQRYLTAFTVSE-----HGQLQGM 544
OY 635 KYIILGSTRVPAHKKPRLLYNGELICOTOSGKVNNTIYSTHGFSLNKDXGPE----- 688
    || || || || || || || || || || || || || || || || || || || || || ||
Db 545 -----OQKNLHCGKLV-----SVSPNPFY-----VRSKGWDEEDNRDE 578
OY 689 -----LRPMLAHN-----LMLKRFSDAMEMKILNDEAMN-- 719
    || || || || || || || || || || || || || || || || || || || || || ||
Db 579 RTIGKTLVAKCLREFLGNENCDESTRKAMMDSFYLTISMAAKRAIOFISDSWDIM 638
OY 720 -----ELARACLHMEVEFAIRYRRIGNVGIWLSBOIGIEDNYNLACHLAMF 769
    || || || || || || || || || || || || || || || || || || || || || ||
Db 639 ASMSIKTRLDYAVMCLGIMKNVGRABRAVRSQNGENDSMCALATELSLWEALITLY 698
OY 770 TND--YNAADLYLASSCPITAALEMRRDLOHWSALQAKHLAPOIPITS--EYAIOL 825
    || || || || || || || || || || || || || || || || || || || || || ||
Db 699 AONERYDMLNKLQOSOM-----WSSAFELAE--TKDRILHRLNHYNAKYL 743
OY 826 BEAGYVNV---ALAHYKGITGDNKEHDEACLAGYAOMSIRNGDLIRGVNQLAKHPSPVL 882
    || || || || || || || || || || || || || || || || || || || || || ||
Db 744 EARRQASTEALEIENYER-----AGVHAEEV-----FRRL 773
OY 883 KRDCGAILENMKQFSEAOLYEKGLYDKAASVYIRSKWMAKV-GDLLPHVSSPKTIHQY 941
    || || || || || || || || || || || || || || || || || || || || || ||
Db 774 KD-----YPKQIEQYVRRKREERYSKILKNINSQAFRSLEY 809

```

```

OY 942 A---KAKEADGKYKAEVAAYENAKOMQSVIRIYDILHNNPEKAVNIVRETOSLDGAKVA 998
    || || || || || || || || || || || || || || || || || || || || || ||
Db 810 SWMGAYLESVGELEGALEIFYSYSAKDYCMVRVKICQ-GKTDEARLAEESKDKACYLLG 868
OY 999 REFLOLQGSALIOFLVMSKCNNEAFITLAAQCHNKKEITADIT-----GS-----EDT 1045
    || || || || || || || || || || || || || || || || || || || || || ||
Db 869 RMYENDGDVYKAVKFFTKARALLSAILRAKEHDMKRLANCLMAGSGELVSARYEDL 928
OY 1046 TNEGYOSIALY-----FEGEK 1061
    || || || || || || || || || || || || || || || || || || || || || ||
Db 929 PCYAKRAWLHYKACMIGRALDLARTQFSALDLITDLDAGTDPKILRRAREFFENQ 988
OY 1062 RYLQAKKEFFLLCGYSRALKHEFKCPSSEDNVAIEMAITVGOANDLITNQLDLHLLGE 1121
    || || || || || || || || || || || || || || || || || || || || || ||
Db 989 NYEKAVNFECLAKEPSGAVQ--LCNNRNVRSDFKA-----ELMT-----PT 1028
OY 1122 NDGMKDKAYLEFLMALKQVREAOQTIIAREE-----QSGNRYNADVLFSYAEK 1177
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1029 KDDMPN-----VOERQVLETVAEELCLOQAYSAAKKFTQAGDKLSARALLK 1077
OY 1178 S---QKIRIPSEMATNMLILHSYIIVKIHVKNGDHMKGARMLIRVANNISKPPS--HIVP 1232
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1078 SCDIKIRFPANTARKKEI---YILANPLQTTMQDNOQTFMKDIEFTYTSQSFEHLGN 1134
OY 1233 ILTS-TVIEC--HRAGLKS--AFSPAM-LMRPEYRSK-----IDA-----KYKKKT-- 1274
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1135 FYKSAVIAIEAEHLRTLDRSMGALEMAACVLEAEOKNMSTGLDALKBDLKKYVQLRKL 1194
OY 1275 -----EGWYRPDPIS--IEEATTPC 1293
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1195 QIMKDAADGMKQRLITTLAEESIDDIIVPC 1223

```

RESULT 8
 T15842
 Hypoetical protein C54G7.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T15842
 R:Du, Z.
 Submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid C54G7.
 A:Reference number: Z18416
 A:Accession: T15842
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1119 <DZ>
 A:Cross-references: EMBL:U0410; NID:g1065453; PID:g1065457; PIDW:AAA81394.1; CESP:C5
 C:Genetics:
 A:Gene: CESP:C54G7.4
 A:Introns: 6/3; 70/1; 111/2; 160/3; 180/1; 277/1; 325/3; 367/3; 395/3; 430/3; 494/1;

Query Match 2.6%; Score 190.5; DB 2; Length 1119;
 Best Local Similarity 16.8%; Pred. No. 0.00087;
 Matches 244; Conservative 227; Mismatches 500; Indels 483; Gaps 69;

```

OY 46 SILEKTTWLGAP-----IOPAWOKTSGNYLAVTGADYIVKIFPRHGQKRSKSEINLPNCVAMD 101
    || || || || || || || || || || || || || || || || || || || || || ||
Db 2 SVFRKFNGLPEHGOHLHFAEMWYNSNYIACGALGALKVY-KIGTDATDLNKSPPAAATL- 59
OY 102 WDKDGDVLAIVAEKSSCIYL--WDANTNKTSLD--NGMRDMSFLMSKYGSFLAVGTV 157
    || || || || || || || || || || || || || || || || || || || || || ||
Db 60 -----VVOALAECHNATVMATVMENNQKLTSPSTGL-----TIWVG----- 97
OY 158 KGNLXIYHQTSRKIPVLGKHTKRITCGCMAENL-XALGEDKMITVSNOEGDITRQTO 216
    || || || || || || || || || || || || || || || || || || || || || ||
Db 98 -----MFEQWCEEM--INNRNSVAVSICWNLDEGTKIIAIAADGVAVIGTLBGNRIWKE 151
OY 217 VRSPEXNNQOFELMKDDDTSAESW--ISVVLGKKTFLFLNLNEPDNPADLEFOODFGNIY 275
    || || || || || || || || || || || || || || || || || || || || || ||
Db 152 L-----EIQLAACELDIDMHCLAEADLEQALAKK-----EHQKE--EIV 188

```


Oy	276	CUNYAGGRIMIGSCHFUVVISTHNGELGQELFOQRANHKOLUTSLAVSOTLKNVATCGD	335
Db	189	CLKTW-----	197
Oy	336	NCIOTODVLKDMVYVILNDEENKIGTSLMSTDGOLLALSTORGSLHFVETK-----	389
Db	198	---KSKTILIDEMENDFERKLEKREBYTGTAIFLN-----	238
Oy	390	-----LPIIGDACSRIYALUTSLEEVYANVEGELITVSDVENPFAVAGLYH	439
Db	239	PEHEKNEPYQVPDPBRFFVAVAYARGMMQ---LMRSLNDPEEVVSI---PRFKITG---	290
Oy	440	LAVOMNNRANFYVLGENAVK-----KJADMELG-----	478
Db	291	-AKMSPGALFALVCGNDSKDDEPTFSKIHFLSAYGHIVGFPOTNDSCTJTGICESTGJLRM	349
Oy	479	---YAAALFEGKVOJLH-----ISELIDAOEERETRLPAY-----DDK-CRICHJLT	524
Db	350	AIAMAAGTLLIGQIRPERPKMSIETIYVYVYOKELVOTGIMYDYKTDEKTYKTYTHEN	409
Oy	525	SDF-----LIYGTGNGVOYFIEDMOFVN---DYRHPVSYKJIFPPNGTRLVFIDEK	575
Db	410	MAFYREHCVLINQDQDGVIRHYFCQJLNSIGTSLDY-NTYTVRRPFACTNGCAITASE-	467
Oy	576	SDGVYCPVNDATYEIIPDSSPTIKGYVLMENPMQKCVFIAYODDKVYTVYFHHKDTIGAK	635
Db	468	DYRFIWM-----HFVLPKPFMSVOAGI---HVPEKSGDYVLEEQORTIEY-----GTK	510
Oy	636	VILGSTRVPFAKHPRLLYNGELTCTGTOSQKVNIYLSYHGLSLNKKQGPDELPMJLH	695
Db	511	RJLDSKOEI-----CALCIG-----	525
Oy	696	NLMKRPDSAMEMCRILINDEAANNELARACLHHEVEFAIRYRRIGVYMSLEQIKG	755
Db	526	-----DFFPMALLSG-----GJYRSLNDGVI-----	548
Oy	756	IEDYNLLAGHLAMFTNDYNIJLAODL-YLASSCPJALAEHRDLOHMSDALOLAKHLAPQI	814
Db	549	-----TNSYPVSPINSIKINCTFTRL-----AVIKIV-----EQV	579
Oy	815	PEISKEVAIOLEFAGOVVNMJLHYEKGJTGDNKEHDEACLAGVAMSTIMGDIRRGVNO-	873
Db	580	PFQFLY---ERDGDCLKLI-----YTSDKRD-----VWQYEMDQNN5	614
Oy	874	---ALKHPSRYLKRDCALILENNKQFSEAAQIETKGLYDYKAA5YIRSKNNAKVGDDL	930
Db	615	NMLALTKOKKILIDGNSILE-QSSVNGSILFLQ-----NLVYRANVEIKI---LLT	662
Oy	931	HVSSPKIHLQYAKAKREADGRYKRAVYAYENAKOMQ5YIRIYLDHJLNPEKAVNIYRETO5	990
Db	663	PENFTKCIIVEVMIKA-----KQVNMJLVNAML-----EADIDYMERSPH	703
Oy	991	LDGAKNVARF-----FLOLGOTGSAJQFLV-----	1016
Db	704	SELSMSTIANAVEFKHAFCDEHAFVYKIGDV-AGVQDLAVEMKKVGDYASVYDLIRNDD	762
Oy	1017	---SKCNNEAF-TLAAQHNKKEIYADIIG5-----EDTJNE-----DYOSIATYFE	10588
Db	763	DEGCG--DARRNAEHTHEEMEMBEAQAAY5YCGDTIRANOIDLINGNLFGELBYLARL5	820
Oy	1059	GKRYLOYA-GKFFLLGQY5RALKHFLKCPSSDVAIEMAIETVGOAKDELITJQOLIDH	1117
Db	821	DQSEFMEWGAFAF5SRGCMQDAVECYLR--RSLPQKALHACKELNOMQKAPFIAD--ANH	876
Oy	1118	LGEGDQMPRDAKTLEFLYALAKQYREAPQATJIIIRQ5Q5AGNTR--NAHVULE5SMYAE	11757
Db	877	M-----ENIADNRK5YVPELKCQCYVLG--AVLVENRHQOTIKELRKKIDKNHFE DALD	930
Oy	1176	---LKSOKIKIPSEMATNMLI5YILVIKIHVNGCHMGAMJLIRVANISIKF--PSHI	12300
Db	931	ESGLASQDSRILENTYMWGALAFHFHMLMAQOHPFE-KRIDALQTL5LSDYEELDPRAEV	989
Oy	1231	VPILSTVIECHRAGLKNSAF5FAPAMLMREYRSKIDAKYKKKIEGMWR-----PDISE	1285

Db	Accession	Score	DB 2	Length	1124
Db	990 HSMILAAANNQGFICSKA-----MMRLAEFEFFDDAEKEEMNNLSFRLPSENPYPNP	1043			
QY	1286 IEEATTPCPFCFLLPECELLCPGCCNSIPYCIATGRHMLDDMTVCPCDFPALYSELK	1345			
Db	1044 -NSAAVACSVQDAKIDPDLQCSCEQCFKFPYCIASGRLLDNIFFWLCPCRCRHHQHQRIP	1102			
QY	1346 IMLNTESTCPMCSE 1359				
Db	1103 ---KYNCCPLCHD 1112				
RESULT	9				
H88772	Protein F23B2.4 [imported] - Caenorhabditis elegans				
C:Species:	Caenorhabditis elegans				
C:Date:	10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001				
C:Accession:	H88772				
R:anonymous:	The C. elegans Sequencing Consortium.				
Science 282,	2012-2018, 1998				
A:Title:	Genome sequence of the nematode C. elegans: a platform for investigating bio				
A:Reference number:	A75000; MUID:99069613; PMID:9851916				
A:Note:	see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_elegans/				
A:Accession:	H88772				
A:status:	preliminary				
A:Molecule type:	DNA				
A:Cross-references:	GB:chr_IV; PID:CAB05177.1; PID:g3876274; GSPDB:GN00022; CESP:F2323				
A:Gene:	F23B2.4				
A:Map position:	4				
Query Match	2.6%; Score 187; DB 2; Length 1124;				
Best Local Similarity	18.9%; Pred. No. 0.0015;				
Matches 198;	Conservative 145; Mismatches 395; Indels 310; Gaps 45;				
QY	76 YIVKIFDRHGKRSEINLPNCV-AMDDMDDDVLAVIAEKSSCIYLDAMTNTKTSOLDN	134			
Db	81 WVDKILD-----ENNEGVCITYDLAKFRPDSSELLAADNK-VYLPDVNEGGMOTLK	131			
QY	135 GNRDQMSFLMSKVSFLAVGVKGNLXIYN-HQTSRKIPVGLKHTKRI-----	182			
Db	132 GHKDLVYTVAMSHNGELFASGGADKLVIIMNEKHGTLRY-----SHTDVICOMFENPCNQ	187			
QY	183 -----TCGC-WNAE-NLXALGEGDKMIV-----SNQ	207			
Db	188 ILLTALNEFGLMSTADKNVTKORSVVRCCSCAMWTDGTFIIGHDGTITLRKTKNATE	247			
QY	208 EGDTRKOTQVRSEP-----XNQPFLLKMDRPTSAESMTISVYLKGTLEFFLN---	257			
Db	248 EESIIQDR-NEPILWGIAFSSNRKTFASRDSGNGMIDEIMAVIDNNKTKLSFYLDGTF	305			
QY	258 -EPDNPADLEFQODEGNIVCYNMYGDRIMIGFSGHFVIVISTHTGELGOELFOARNHKD	316			
Db	306 IESKN---LEFEPH-----CISYCLNGEYLLIGGSDKIITKTYRGRGVLGTVAGQDHIMW	357			
QY	317 NLTSAVSQTLKAVATCGDNCIKIDLDVLDKMYIIMDENRKGILGTLMSDQGLAL	376			
Db	358 SYTVRPNSTV--AMGCVDTGTACYNLV-----FSTVHCVDHAR---Y	395			
QY	377 STORSLHVFLEKLKP--ILGDACSTRIAYITFLLEVTVANPEGELPI-----	422			
Db	396 ANRKSMTDVFYONLEKRRSSNICHDLYKKMSLYTKTALAVQLSDKITQIKOTGVSKNER	455			
QY	423 -----TVSDVEPNFAVAGLYHLAV-----GMNRRAMFYVGENAVKK	460			
Db	456 RKQLYTLQDTIRKDLSELMVYTHGHLVVCNDEKTECYDFGKIKRSM-----N	505			
QY	461 LKD-MEYLGTASICLHSDYAAALFEGKVQQLHLESEILLDAQEEETRLFRAVDKCRIL	519			
Db	506 MKSIVRYL-----RVLGPAHRET-----	524			

```

QY 520 CHALTSPFLIIG-IDTGVVQVFFYIEDMOQFVNDYRHPVSVKKTFFDPDNGTRFLVFIIDEKSDG 578
Db 525 -----LVLTGTDGGVYKVFIDNDYPIILDSR-KTAIKCIDINANRTVLASIEDF--- 572
QY 579 FVYCPVND-AVVEYIPDFSPPTIKGVLMWENPMDKGVTAIVDDKVVYVVFHKDTLOGKVTI 637
Db 573 -LVCKMSDIATGEIILLQEPGCSVVFNTVYENMLPFTTNMMLHVRFLAARFGHTTRGVYV 631
QY 638 LAGSTKVPFAHRPLLNLGELTQTQSGKVNNIYLSTHGFSLMKDXGPDELPMLAHNL 697
Db 632 LGFYKNRTFC---LVQYN-----LIPLEV-PYTIHLX 659
QY 698 MKRFSDAMEWECRL-----NDEAAMNELARACLIHMEVEFAIRYVRIGN---VGIYM 748
Db 660 QYIERGQPEKRLRIACLGAVKND--WKYLANKALDLEFDVARKAYKVRDRKRMKRMW 716
QY 749 SLEQIK--GIIDVYMLAGLHMFNDYNIADODLYLASSCPIALERDRDLOHMSALOLA 806
Db 717 ELKMKSGNGEPDAILRATILIA-YTKKFPREAKIFKNGENRAMELFETDMRMDDVQE-- 773
QY 807 KHLAPDOIPIFSKEYAIOLEFRAGDYVNALAHYEKGTGONKEHDEACLAGVAQMSIRMGD 866
Db 774 -----VMTTASGETKMKMLRRKASWARDANOPKIA-----AEMLISSGD 812
QY 867 IRRG-----VNOALKHPSRLKRDGCALITLNNK-----QPSAQAQIYKGL 907
Db 813 LDKRAALIINDMIELATELAIISHKIDRSD---LETMKKLSAIFIRKHEGLASRIPOSIN 868
QY 908 YDRAASVYIRSKNWA---KVGDLLPHVSSPKIHLQYAKAKREADGRYKEAVVAENAKQW 964
Db 869 DKMSIVDMHVAGHMTDAFADLRHPKYVE-DVLYPYANFLERDRFEEDQKAFHNRGKE 927
QY 965 QSVIRITLHDHNNPEKAVNIYRETQSLD 992
Db 928 QEAMHV-LEQLTS-----NSVNNRFPAD 949

RESULT 10
A36349
clathrin heavy chain 1 - yeast (Saccharomyces cerevisiae)
NAlternate names: protein G165; protein YGL206c
CSpecies: Saccharomyces cerevisiae
CDate: 01-Feb-1991 #sequence_revision 01-Feb-1991 #text_change 20-Jun-2000
CAccession: A36349; S64225
Rilemon, S.K.; Pelliscena-Palle, A.; Conley, K.; Freund, C.L.
J. Cell Biol. 112, 65-80, 1991
A:Title: Sequence of the clathrin heavy chain from Saccharomyces cerevisiae and requirem
A:Reference number: A36349; MUID:91093343
A:Accession: A36349
A:Molecule type: DNA
A:Residues: 1-1653 <LEM>
A:Cross-references: GB:A52900; NID:g3535; PIDN:CAA37082.1; PID:g3536
R:Freuenmann, M.; Potter, S.; Souciet, J.L.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64044
A:Accession: S64225
A:Molecule type: DNA
A:Residues: 1-1653 <REU>
A:Cross-references: EMBL:Z72728; NID:g1322841; PIDN:CAA96919.1; PID:g1322842; MIPS:YGL20
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:CHC1
A:Cross-references: SGD:S0003174; MIPS:YGL206c
A:Map position: 7L
C:Superfamily: clathrin heavy chain

Query Match 2.5%; Score 182.5; DB 2; Length 1653;
Best Local Similarity 18.4%; Pred. No. 0.0052;
Matches 288; Conservative 222; Mismatches 524; Indels 535; Gaps 75;
12 LGDGNNAVACLETGTCACVLIHSSRSGSSVE---MKRIFSL-----LEKWTGAPIOFAWQ 62

```

Dd	54	LAKGEVTRKMMGGDSALIMPSQWISYVRANGTIVQIFNLETYSKLSFLLDEPVIF--WR	112
Oy	63	KTSGNYLAVTGADEV--KIFPRGQRKRSSEINLPQNCVAMDMDKGDVYLVAIEKSSCIY	120
Dd	113	WLSEITTEGFEVTRARSILTSNVFD-----GN-----VNAKQDILL	145
Oy	121	LMDANTKTSQLDNGMRDQMSFLM--SKVGSFLLAVGVTRKGNLHYNHQTSRKPIVLKHT	179
Dd	146	LRAHALNMT-QIINEVAKRN--LDMFAVVGGLDENGRIADRIOLFSQO--RMTS-----	194
Oy	180	KRITGGCNAENLKAALGEDKMITVSNQOEGTLINQTOVRSEPNXMOEFLKMDRTSAE	239
Dd	195	-----QAIDSHAIFNILLTEGNGSTPRVOY-----FVTSNRNATTCAGE	233
Oy	240	SMISVY-----LGKKT--LEFLNLNEPDNPADEFOODGEGNIVCYMWGDRIMIG	288
Dd	234	LRIEIDHDASLPSQOYKETTDETFPRPDATNDPIAVQVSEKGIYILLTKYG-----	286
Oy	289	FSCGHFVVISHTGELGCEITQAR-----NHKNL-----TSI	321
Dd	287	-----FIHLYELETGTNLFVNRTIAESVETAPARYNHENGIIACINKKGVLAVEISTSQ	339
Oy	322	AVSOTLK-----VATGDCMCITIODVLDKMWVILNIDEEN-----KG	361
Dd	340	IVPYILNLSVALALIAVATFG--GLPCADDLFOKQFESLLQNDYQAAKVAASSTSLRN	388
Oy	362	LGTLSWTDQO-----LALAST--QKQSLH--VELTKPLIGD-----	395
Dd	399	QNTINRLKNIDAPGAGLSPILLYRSTLLDKGLNKKEETIELAR-PVLOQORQOLFEXWLK	457
Oy	396	----ACSTRIAYLTSLEEVYA-----NPBGELPIVSVDEPN	431
Dd	458	EDKLECSSEELDIAKPEPTTLTACLYRAGAHAKVISCIAELDOEFKILIPYQCKQVGYOP	517
Oy	432	FVAVGLYHLAVGMNNRMFAFYLGS--EVA-----YKKLDM-----EYLGTVASICLHSD	478
Dd	518	FLVL--ISSLRSSPDRSEFPAVSLONPETRASQIDIEKIADLFPFSONHHQGSTSLLDAL	576
Oy	479	YAAALFEKQVOLHLIESEILDQOE-----ERETRLF	509
Dd	577	KGPRDQGHLOTRLVENVLHLAPQVADAILGNNTFISHYDKPTIASLSEKAGLQORALEVN	636
Oy	510	PAVDD--KCRILCHALTSDELIT-YGTDTGVYQVF-----YEDMOFVN	549
Dd	637	TDIKIRKCVVHTNAPLIDMLVIGFGKLANEOSLACLKALMDNNINQANTQVQVATKFS	696
Oy	550	DYRHHVSVKKTFFPRPNGR-----LVFIDEKSDGFVCYPNDAITYEIDRSPITIKG	601
Dd	697	DLIGSTLIKFEEDYNATEGLIYVLASLVNLTBCKD--VVIKYL--EAAAKKQYR--ELERI	753
Oy	602	LWENMPMDGVFIAYDDKQVTVYFHHND--ITQAKVILASTSTVFPFAHKBL--LYNGEIL	659
Dd	754	VKDN-----NYDEREYKVNFL--KDNALMEDOLPLVICYDRFVYHEKILILYKTSQ--	801
Oy	660	QOTSGKAVNNIYLSHTGFLSLMKDXGDELRYMLAHNLMLKFSDAWEMKRIILDEAAMN	719
Dd	802	-----NLKEFIYU-----VQOVNPSKTAQVUGALLDMD-----C--DEAFIQ	836
Oy	720	ELARACLIHMEV--EFAIRVYIRIGNVQIWSLEO--IKGIED--YNLLA-----	763
Dd	837	SLLOSVLGOVPINELTTEVERKRNRLKILLPLEESLSSOGIODAVUNALAKIYIDSNSNP	896
Oy	764	-----GH-----LAMEFNVDNL--ADDVLASS	784
Dd	897	EKFLKENDQVPTLDVGNCEKRDYULAYLAEKGQDNDLIRITNESMKYUARYL---	953
Oy	785	CPIALEMRRDLOQWDSAL-----OLAKHLAPDQIFPISKEVATOLEFFAGDYVNAL--	835
Dd	954	-----LE-RSDLDLNNKVLNENHRRQLLDSIVSVEIPELTDBEPYSLVQVAFMNGKL	100
Oy	836	----AHYEKGITGNKEHDEACLAGVAMQSTRMGDIRRGVNOALKH--PSVYULKRDGCAIL	890

Db 1008 LELIELLEKILBSPSPENVALQGLLLLS-----AIKYEPTKV-----SSYL 1050
 QY 891 ENMKOF-----SEAOLYKGLVYDKAASYIRS-KMKAVGDLPHVSS 934
 Db 1051 EKLNDVADADIAPLCIEHDLKEEAFELYDKHEMCKALKVLIEDIMSLDPAASYADKINT 1110
 QY 935 PKIHLQYAKAKKADG-RYKEAVVAAYENAKOMQSVIRIYLDHLNPEKAVNIVRETOSLDG 993
 Db 1111 PELMSQIGTA-QLDGLRIPALIESY-----IKADPSNYENVIDIAE-----1151
 QY 994 AKVAFRELLQDYGSAIOFLVMS-----KCNNEAFTLAQOHKME-----1 1035
 Db 1152 -----QAGRYEELIPFLPMARKTLKPEKIDGALLIAYAEINKHEIENLAGSNVA 1202
 QY 1036 YADIGSEDTNEDYOSIALYFEGEKRYLOAGFFLLCGYSRALKHFLCSPSEEDVAI 1095
 Db 1203 NLHVGDKEFENKRYKARLCYSANVSKLASTLYLGYQAADTARA---SIRKV 1258
 QY 1096 EMAIEYGAQKDELFTNQL-----IDHLLG--ENDGMPKDAKYFRLYMALKQY 1142
 Db 1259 WKLVNDACIEKKEFKLAQIGLNLVHAELDELVERESNGYFEELISLFEAGLGERA 1318
 QY 1143 REAAQTAIITAREBQASAGNYRNAHDVLESMYAEKLS-----QKIDP-----SE 1186
 Db 1319 HMGCFTELAI-----LYSKYEPDKTFEHLKLFMSRINIKVIRAVEQAH 1362
 QY 1187 MATNLMLHSY-----ILVKIYKNGDHHKKGARMLRVAN-----NISKFP 1227
 Db 1363 LMSLVFLVAHYDEMNAALTLEKSTKIDHAYFEKEVAVKSULEIYYKAINFEYKFP 1422
 QY 1228 SHIVPILTS 1236
 Db 1423 SLVLDLTS 1431

RESULT 11
 T21306
 hypothetical protein F23B2.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 R:McMurray, A.
 C:Accession: T21306
 A:Reference number: 219403
 A:Accession: T21306
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1047 <M1>
 A:Cross-references: EMBL:282266; PIDN:CAB05177.2; GSPDB:GN00022; CESP:F23B2.4
 C:Genetics:
 A:Gene: CESP:F23B2.4
 A:Map position: 4
 A:introns: 16/2; 55/1; 95/2; 168/3; 266/1; 321/2; 399/3; 437/3; 591/2; 642/3; 695/3; 751/2

Query Match 2.5%; Score 178; DB 2; Length 1047;
 Best Local Similarity 18.8%; Pred. No. 0.005;
 Matches 191; Conservative 140; Mismatches 385; Indels 302; Gaps 43;

QY 105 DGVLAIVIAKSSCIYLMADANTNKTSQLDNGMRDQMSFLMSKVSFLAVGTVAGNLXIV 164
 Db 27 DGSLLLAADNK--VYLFVDNEGQKQTLKGKHDVLYTVAMSHNGELFASGADKLVIYLM 84
 QY 165 N--HQISRKIPVLGKTKTKRI-----TC 184
 Db 85 NEKHEGLTRY-----SHTDVLCQMFNFCNOILLTLCALNEFGMLSTADKNVIKQSVYRCC 140
 QY 185 GC-WNAE-NLXALGDEPKMTV-----SNOEGTIROTQYRSE-----XNNQFLMK 230
 Db 141 SCAMNDGTGTFALIGHGGTTLTKRKNATPEPSITIQD--NEEIMGAIASSNRTFASRD 198
 QY 231 MDRTSAAESMISVVLGKTLFLINLN-----EPDNPADLEFQDFGNIVCYNMVGDGRIM 286

Db 199 SOGNPMGIDEIMAVIDMNKTLSPYSLDGFIESKN--LEFEHH-----CISYCLNGEYL 250
 QY 287 IGFSCGHFVYISTHTELGCEITFOARNHNDLTSIAVSOTLKNVAMGDCNCKITDLDVL 346
 Db 251 LIGSDKILKIYTRKCVLLGTVAQMDHMTWSVYLRPNQTV--AMGCVDTGTCACMLV-- 306
 QY 347 KDMYVILNIDENKKGIGTSMWTDGQLALSTQSGSLHVFLEKLP--IIGDASTRYVL 404
 Db 307 -----FSTYHCVDHAR--YANKSMYDFVONLEYRTSSNICDLYVK 348
 QY 405 TSLLEVTVANPVEGELPI-----TVSDVEPNFVAAGLYHLAV 442
 Db 349 MSLYDTRKLAQVSDKIQIYQGTGYSKNERKQKLTLDITIKKLSFLMYVTHGHLYV 408
 QY 443 -----GMNBNPFYVLGENNAVKKLD-MEYLGTVASICLHSDYAAALFEGKYV 489
 Db 409 CNDEKLECYDFKGIKRSM-----NMKSIYRL-----436
 QY 490 LHLIESPILDAQERETRLFPVAVDCKRILCHALTSDFILYG--TDPIGVQYFYIEDMOFY 548
 Db 437 -----RVLGGPAHRET-----LVLGTTDGGVYKVFIDNDPIL 469
 QY 549 NDYRHPVSVKIFPPDNGTRFLVEIDKSDGFVYCPYND-ATYELIPDSPTIKGLWENMP 607
 Db 470 LDSR-KTAIKCIDIMANRTYLAISIDPT---LVCKMSDIATGFTLLQEPGCVSVENVTN 524
 QY 608 MDKGVPIAIDDDKVYVYVHFHKDITQCAKYLACSTVPRAHKRLLYNGELTCOTQSGKY 667
 Db 525 ENLFAFTTNMMLHVRLAAGHTTRGVGLGVKKNRTFC--LVQYN-----569
 QY 668 NNTYLSHGLSLNKKXGPELRLMLAHNLMKRFSPAWEMCRIL-----NDEAMANEL 721
 Db 570 -----LIPLEV-PTYIHLQYIERGDFKALRIACLVYKND--WKYL 609
 QY 722 ARACLHMEVEFAIRVYRIGN--VGIVNSLBOIK--GIEDYNLAGHLMFTNDYNLA 776
 Db 610 ANKALDLEFDVARKAYKVRDRCKMLRMWELKMKMSNGSPDAILRATILA-YTKKFREA 668
 QY 777 ODLYLASSCPIALEMRRDLQHMDSALQAKHLAPDQIPITSKEVYIOLFACDYVNAIA 836
 Db 669 AKIFKENGFEENRAMELFTDMRFDDVOE-----VMTASGETKKMLM 710
 QY 837 HYKGITGDNKKEHDEACLAGVAQMSIRMGDIRG-----VNOALKRPSVLRDGC 887
 Db 711 KRASWARDANQPKIA-----AEMLISSGDLKRAALLIIDNMLDELATISHKIDRS 763
 QY 888 ALLENKK-----QFSEAOLYKGLVYDKAASVYIRSKNMA--KVGDLPHVSS 934
 Db 764 --LETMKKLSAYFLRKHEFGLASIRFQSIINDMKSIVDMHVNAGHMWDFAIADRHRYVE 821
 QY 935 PKIHLQYAKAKKADGRYKEAVVAAYENAKOMQSVIRIYLDHLNPEKAVNIVRETOSLD 992
 Db 822 -DVLYPARFLARLDRDREFAOKAFHRAGKEQEAHV-LEQULTS-----NSVNNENREAD 872

RESULT 12
 S51471
 killer toxin insensitive protein 3 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein L3502.7; protein YLR384C
 C:Species: Saccharomyces cerevisiae
 C:Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 21-Jul-2000
 C:Accession: S51471; J05565
 R:Du, Z.
 submitted to the EMBL Data Library, December 1994
 A:Description: The sequence of S. cerevisiae cosmid L3502.
 A:Reference number: S51466
 A:Accession: S51471
 A:Molecule type: DNA
 A:Residues: 1-1349 <D07>
 A:Cross-references: EMBL:U9104; MIPS:YLR384C; NID:g609423; PID:g609430
 R:Yajima, H.; Tokunaga, M.; Nakayama-Murayama, A.; Hishinuma, F.
 Biosci. Biotechnol. Biochem. 61, 704-709, 1997

A>Title: Characterization of IKI1 and IKI3 genes conferring pgk1 killer sensitivity on S
A:Reference number: JC5564; MIDID:97290891
A:Accession: J05565
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1349 <Y>A>
A:Cross-References: DBJ:D87841; NID:g2116969; PIDN:BA20120.1; PID:g2116970
C:Genetics:
A:Gene: SGD:IKI3
A:Cross-References: MIPS:YLR384c
A:Map position: 12R
C:Function:
A:Description: confers sensitivity to killer toxin
C:Keywords: transmembrane protein
F:97-113/Domain: transmembrane #status predicted <TM>

Query Match 2.4%; Score 177; DB 2; Length 1349;
Best Local Similarity 18.2%; Pred. No. 0.0085;
Matches 245; Conservative 188; Mismatches 384; Indels 530; Gaps 66;

```
10 GLLGD---GNAVACLVETGCACVLIHSSRGSSVEMKRIFLEKTLGAPIDQAMQKTSQ 66
Db 223 GLVGNQLRPTMYMDTGVTALDSH-----ITISW-RGDC 259

Qy 67 NYLAVTGADYI-----VKIFDRHQKRSKSEIN-LPGNCVANDMDKGDVLAV 111
Db 260 DYPAVSVEEVPDEDETKSIRRRARVFESREGQLDSASBPVTGMHQSFKQSGSLIAS 319

Qy 112 IAKSSCIYIMDANTKTSQDLNMGMDQMSFLMSKVGSLAVGTAKGLNLIYHNQTSR 171
Db 320 IQRT-----DLGEEDSDVIVFERNG-----LKHGEFDR 350

Qy 172 IPIGLKHTKRTICGCNNAENLXALGEDKMITVNOEGDITROTQYSEPNXNQFLMK 231
Db 351 LPL-----DEKVESCVNNSN-----EALAVLANR----- 376

Qy 232 DDRTSAAESMISVYLKTLFPLNLNEPDNPADLEFOQDFGNIVCY-NMYGDRIMIGS 290
Db 377 -----IQLTMSKKNHYMLK-----QELVASDISYVKKWHPKEDFLMFS 414

Qy 291 CGHFVVISHTGELGQ---EIFQARNHKNQNLTIAY-SQTLNKKVANTCGDNCKIKIDLD 345
Db 415 DAGFINIVDFAYKMAQGPYLEP-----DNGTSLVVADGRTVNTP----- 454

Qy 346 LKDMVYILNDEENKGLGLTSMDDQLALSTORGLAHV-FLTKPLPLDGACS-TRIA 402
Db 455 -----LATANVPPMYRYRDEPFGNVLDVACSPSNEI 486

Qy 403 YLTSLLEVYVANY-----EGELPTYSVDYEPNFVAVGLYHLAVGMNRMAYVLGE 455
Db 487 YAAINDVYLFAAVPSSIEMKKCKRHPISIVCEPFKSEFT----- 524

Qy 456 NAYKTKLKMEYLG-TVAASICLSHYAALAEKGVQJLHLESLIDAOEERETLPFAVD 514
Db 525 SEVDSIRQVAFINDSIVGLDLD-----NLSRIALDDIQDITQPLTIVEE 572

Qy 515 KCRILCHALTSDF-----LIYGTDTGVV-----QYFYEDW-QFYNDYHHPYSVKIRP 562
Db 573 YDKIV-LLRSDPDYHNIYETRDGYVCOQDAGQLMETIKFPQLYRDRF---VNR--- 623

Qy 563 DPNGRFLVFIDEKSDGFVYCPVNDATYEIPDSPTIKGLVEMWPMDKGVFLAYDDKYV 622
Db 624 -----VHNTSAEDDD-----NMSAESSELVAFG----- 646

Qy 623 TYVFKHDTIOGAKVILAGS-----KVFPAH-----KPLLY----- 654
Db 647 --ITNNGKLFANQVLASATSLSEITDSFLLTAAQHNIQFVHLNSTDFKPLPLEEGVE 704

Qy 655 -----NGELTCQFQSGVNNIYLTSHQFLSNLKDQXG-----DELRPMLAHNLMKR 701
Db 705 DERVRAIENGSLIVSIVPSKSSVYVLAQTRG---NLEFYIPRIMVLAVER---KNIMAKR 757
```

```
Qy 702 FSDAWEMCR-----ILND-----EAMWELAR-----ACLHMEVEFAIR 736
Db 758 YKEAFIVCRTHRIINDLIDHYAPELFLENLEAVFINQIGRVDLNLFISCL-SEDDVTKT 815

Qy 737 YVRRIQVGIWLSLEDKIEDIYNLAGHLAM-----FTNDYNLAQD-LYLASSCP- 786
Db 816 KYKETLYSGISKSF---GMEPAPIEMQIYKKKMFDRKTSKVKTCIAVNVLLSNPE 871

Qy 787 -----IAALEMRDLDHMSALQALAKHLAPDQIFPISKEYAI-QLEFAGD-----YVN 833
Db 872 YKKKYLQTIITAVASQNPQNLASALKLISELSENE---EKSCVTYLCFLQDVAVVYKS 927

Qy 834 ALAHYKGTIGDNKEHDEACLAGVAQMSIR-----MGDIRGVNQ 873
Db 928 ALSIVDVSLA-----LLVAQKSQMDPREYLPFLQELQDNEPLRRKFLIDYDYGNEYK 979

Qy 874 ALKHPFRVLKRDQGAILEMMKQFSEAAQLYEGL---YVD--KAASVYRSMNMKVGDL 928
Db 980 ALBHLSEIDKD--GNVSEVYIDYVESHDLYKHGLALYRDSKQVYI---NIYAK----- 1030

Qy 929 LPHVSSPKIHLQYAKAKREADGRYKEAVVAYENAKQWQSVIRIYLDLNNPERAVNIRET 988
Db 1031 --HLSSNQYTTDAAYAYEMLGKLEAMGAYQSAKRKRREMSIAVQKF--PEEVESAEBL 1086

Qy 989 QSLDGAKMYARFFLLQDYGSAIQFLVMSKCNNEAFTLAQHNKKEIYADT-IGSEDTTN 1047
Db 1087 IS-----SLTFEHRYYDA-ADIOLEYLDNWK 1111

Qy 1048 EDYOSIALYFEGEKRYLQAGKFFLLQGYSRALKHFLKCPSSSDNVVAIEMAIETYQAKD 1107
Db 1112 E--AVALT-----CKARVYDASLVATKA---KRD 1136

Qy 1108 ELLTNQDLHLGENDGMKRD--AKYLFRLYMALKQYREAAQTAIITAREOSAGNYRNA 1165
Db 1137 ELL-EVVDPGGLGEGGFIYLAELLADCKQINSQRLRLE-----LRAKKEENP----- 1183

Qy 1166 HDVLSMYAEKLSQKIT--PSEMAT 1189
Db 1184 ---YAFYQETEQADDVSVAPSETST 1206
```

RESULT 13
T17224
Hypothetical protein DKFp434B165.1 - human (fragments)
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17224
R:Dueterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: Z18727
A:Accession: T17224
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-327;328-558 <DUE>
A:Cross-References: EMBL:AL117421
A:Experimental source: adult testis; clone DKFZp434B165
A>Note: the cDNA sequence contains a -1 frameshift near codon 327
C:Genetics:
A>Note: DKFZp434B165.1

Query Match 2.2%; Score 161.5; DB 2; Length 558;
Best Local Similarity 22.6%; Pred. No. 0.021;
Matches 91; Conservative 47; Mismatches 148; Indels 117; Gaps 13;

```
Qy 767 AAFNDYINLAQDLYASSCPPIALEMRDLQHWDSALQALAKHLAPDQIFPISKEY----- 821
Db 11 ALEKDFQKAEGLILRAQPPGLALNYKKGGLMSDALRCKYVPSQLEALQEEYEREAT 70

Qy 822 -----AQLPEFAGDYVALAHYERKGTGDNKEHDEACLAGVAQMSIMGDIR 868
Db 71 KKGARVEGFVQDARHMEAGETSRVAVCYLKVRSNGSLAEKCMKMAELSTIFLPPQ 130
```

```

OY 869 RGVNALKHPRSRYLKRDCCAILLENMKOFSEAAOLYEKGLYDRAKASVYIRSKNMKAVGDL 928
DB 131 RMNEVLA-----VSPOLIGIKHSAAMELYLNLDLVNEALDAFLEGEEMNK----- 177
OY 929 LPHVSSPKTHLOAYK--AKADGRKAEAVAYENAKOMOSVIRIYLDHLNPEKAVNIYR 986
DB 178 -----AKRVAKELDPREYEDVDH-----YKEFLKNNGK----- 206
OY 987 ENQSLDGAQKVAR--PFIOLDGYSAIO-----FLVMKCKNNEFT 1025
DB 207 -VDSLGVADVIALDLVYHOGQMDKCTETATYKONKILHKVYALVATHLIRGSSAQALA 265
OY 1026 LAQOHNK-----MEIYADIGSEDDT--NEDYOSIALYEEGKRYLOAGKFFLLCG 1074
DB 266 LVYOHAPRANLQNFNIYKRIFFDMVSSPGTNCALYHSA-----DLRDLVFLVAVL 317
OY 1075 QYRALKHFLKCPSSDENVAIEMAIETVQAKDELITNOLIDH 1117
DB 318 SPSSSVKTVWK--SSEAN-----SPAHEEFKTMILLIAH 348

RESULT 14
117345
hypothetical protein DKFZp586M1824.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17345
R:Duetschehoef, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: 218727
A:Accession: T17345
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1223 <DUE>
A:Cross-References: EMBL:AL117665
A:Experimental source: adult uterus; clone DKFZp586M1824
A:Genetics:
A:Note: DKFZp586M1824.1

Query Match 2.2%; Score 161.5; DB 2; Length 1223;
Best Local Similarity 18.8%; Pred. No. 0.072;
Matches 258; Conservative 173; Mismatches 484; Indels 461; Gaps 64;

OY 85 GQKRSEI--NLGNCYAMMDKGDVLAIVAEKSSCIYLDANTKTSOLDNMRQMSF 142
DB 45 GQNHSHIVNL--CPLOTEDDKQALLSTSMRD--VKCWDIATTECSWTLPSLGGFAVS 99
OY 143 LMSKV--GSFLAVG-----TVGNLXIYNHOTSRKIPVLRKTKRITGCGWN 188
DB 100 LAFSSVDIIS--LAIGVGDGMIRWMNLTSTKNNVDVKNFNOGVK-----SKVTALCMH 150
OY 189 --AENLXALGEDKMITVSNQESDTRIQTQVRSFXNMQFELMKMDRTSAESMISVVL 246
DB 151 PRKEGCLAGFTDQKV-----GLDYTSNKPPOISSTY 183
OY 247 GKTLFLNLNPNADLEFQODFNGIYCYNMVGDGRIMIGSCNHFVYISHTBELQ 306
DB 184 HKTVVTLAMGPRVPMPSLGGEDRPSLALVSCGGEG-----IVLQHPKPLSG 232
OY 307 ELFOAHNHNDNLTSIAVSQTLNKVATCGDNCTKIQDLVDLKDVMYLLNDEENKGLGTLS 366
DB 233 EAFD-----INKL-----IRDTNSIKYKLPVNT-----EIS 258
OY 367 WTDGQLLALSTQSGSLHWFELTKPLDAGCSTR----- 400
DB 259 WKADGKIMALGNEDGSEIEF--QIPMLKILCTIOQHKLIVNTISMNENGSOPELSYLA 316
OY 401 -----IAVLTSLLEVYANRVGELRITYSVQVERNFPAVAGLXHLAVGNNNAMPYVLG 454
DB 317 SGNNAVITYVHNHKTVESSP--ESPVIT--EPYRTLSG--HVA-KTYSVAW----- 362
OY 455 ENAVKRLKMEYLGTVASICLHSD--YAAALFEGKVOLHLESEILDQERETRLFPAY 512

```

```

DB 363 -----SPHHNDGLVASVDGTAAQ-----VMDAL--KEEPLCNR 394
OY 513 DDKRILCHA---LTSDFLIYGTDTGVVOYFYIEDM--QFVNDYRHPVSKKLPDPNGTR 568
DB 395 GHQGRLLCVAMSPRLPDCIYSGADD--FCVHKWLTSMQHSRP-----PQGRK 440
OY 569 LVFIDEKSDGFYCCP--VNDATYELIDPSPTKGS--VLMEN--WPMQGVFIAND 617
DB 441 STELEKR--LSQPAKAKKKKKPRLTRPVKLESTIDGEEESMKENSGPVNGVSDQEG 497
OY 618 DDK-----VYTVV--FHKDTIO--GAKYILAGSTVPRANRPFLLY 654
DB 498 EQDAREPRLPCGLAPAVSREPIYICPVSSGFESKSKYTINNKYLL--KEPRKEPETLI 555
OY 655 NGE-----LTGOTOSGKVN----- 669
DB 556 KRRKARSLPLSTSDHRSKEELHODCLVLAFAKHSRELNEDVSADEERFHLGFTDRA 615
OY 670 -----IYLSHGFELSLKDXGPDRLPMLAHNLMKRFSDAM--EMCRILNDEAAMNEL- 721
DB 616 TLYRMIDIEGKHLEN-----GHPELPHQML-----MKGDLGVLOTAAERGETL 661
OY 722 -----ARACLHMEVEFAIRYVRIGANGIYMSLEQIKGIEDYMLLAGHLMFTNDYML 775
DB 662 DNLVMAAPAGYHVM--WAVEAFK-----QLCFQDQYKAAHSLIHKHYE- 708
OY 776 ADDLYLASSCPIALEMRDLQHDMSALQALAKHLAPDOIPTISKY--AIOLEFAGDY 832
DB 709 -----AVELLSNHFYREAIYIAKARLREPOVPLKDYLSMGTVLERGHTYA 755
OY 833 NALAHYKGITDGN-----KEHDEACLAGVQMSIRMDIRGVNOALKHPRSYLRKRC 886
DB 756 VAKCTLGATCAVDAKVLAKKGDASLTPRAELAIYVEDELSLALRCAQOELLANN 815
OY 887 GAIENMKOFSEAAQ-----LYEGLYUDKASVYIRSKNMKAVGDLPR 931
DB 816 WYGAQALDLHESLQGRVLFCLLELSRHLNEKQISEKSSSY--HTW--NTGTEGPF 871
OY 932 VSSPKTHLOYAKAKKADGKREAVAYENAKOMOSVIRIYLDHLNPEKAVNIYRETOSL 991
DB 872 VERVTAVMWSISLSDPREDOYQEFQKLNK-----YPSATVNTPAKOLLNHCIDL 923
OY 992 DAKMVARFELDQVGSALQFL--VMSKCNNEATFLAQNNKMEIYADI--GSEDDT 1046
DB 924 -----TLAVLSQWASWDEAVQALLRAVRSYDSGSTIMQ-----EVSATFLPDGCDHLR 974
OY 1047 NE--DYOSIAL--YFEEGKRYLOAGKFFLLCGQYSRALKHFLKCPSSD-----NYA 1094
DB 975 DKGDHQSPATPAFKSLAFLYGRLYERFMSLSR-----PCPNSSVAVRAGHRTLSYE 1028
OY 1095 IEMAITVQAKDELITNO-----LIDHLIGENDGKPRKAKYLF-----RLYALKQ 1141
DB 1029 PSQOLDASTETDPTESQPERNPSELRLRLEEGEERMLSTFKELSEKHSIQLNSQRT 1088
OY 1142 YREAAQTALIIAREEQA-----GNRYRNAHY-----LFSMAELKSOK---IKIPS 1185
DB 1089 VAEVQETLAEMLRQHKQSOLCKSTANGRPKNREYEAEDPLCSGQCKEERKESLPL- 1147
OY 1186 EMATNLMILHSYTVLVKIHVKNGDHMKGARMLIRVANNISKFPSSHIVPILTSVTIEC 1241
DB 1148 -----ELTKRLTEANRMKAFPESTIKAMPFPVPLEC 1178

RESULT 15
T38393
clathrin heavy chain - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T38393
R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: 221791

```

A:Accession: T38393
A:Status: preliminary: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1666 <MCL>
A:Cross-references: EMBL:Z69240; PIDN:CAA93228.1; GSPDB:GN00066; SPDB:SPAC26A3.05
A:Experimental source: strain 97zh-; cosmid 26A3
A:Gene: SPDB:SPAC26A3.05
A:Map position: 1
A:Introns: 14/3; 39/2
C:Superfamily: clathrin heavy chain

Query Match 2.2%; Score 161; DB 2; Length 1666;
Best Local Similarity 18.7%; Pred. No. 0.13;
Matches 251; Conservative 167; Mismatches 445; Indels 476; Gaps 65;

QY 192 LKALGSEDMMITYSNOEGTIRTOVRSPPXNM-OFFLKKMDRTSAESMTS---VYL 246
DB 15 LKASVGIOSSFGFANVTLSDDKYVCVRDNGVNOVIVLEDDPSVLRPRISADSVILH 74
QY 247 GKRTFFELNLEPDNADLEFQDDFG-----NIVCYNNYGDGRIMGF-----SCGHFVY 296
DB 75 PKKTIILAKAQRLOQVFDLEAKAKINSYMNQDVYWTWISDS--YIGWVTDTSVPHMTY 132
QY 297 -----ISTHTGELQET--FOARNHKDNLSTIAVSQTLNKVATCGDNCI-----KI- 340
DB 133 GSGSDPYKMFDRHSSLSNGTOIISKSNYNEMWFLLIGISSRDNRKA--GNLOLYSKKRKYS 190
QY 341 QOLVLDKDMYVIL--NLDEENKGLGTLSTMTDGGOLLALSTO--RGLNVLFTKLPIIG- 394
DB 191 QPLESHASFAVAVIOPEGVHE-----VOVLALASRLPTGS-----KSLIVAY 232
QY 395 DACSTRIVATLSLEVTVANPYEGELPTVSVYDVEPNFVAV-----GLYHL-----AVG 443
DB 233 DKNPNPAPATATVLDLFFPEPAVNDPFIATIGSTYVNAVYVTKGFHIVYDLETAKCIV 292
QY 444 MN---NRAMFYLVGENAVKCLKDMETLVASICLHSDYAALFECKVOLHLESITLDA 500
DB 293 MNRVSGESIFVTAAKSVGLMAINKGVLVSINPE-----TIIPIYLSMLND- 342
QY 501 QEERETRLPAYDDKCRILCHALTSDFLYGTDTGVVOYFYIEDMOFVNDYRHPVSVKKI 560
DB 343 -----PGL--AVRAMSHA-----NLPGADNLVMOOF--QOLMAOGNYSBAKVAAS 384
QY 561 FPDPMGTRLVFIDEKSDGVYCVNDATYEIPDESPTIKGLVLEMMPMDKGVFIAYDDK 620
DB 385 SP-----RGLRLTSQVIDQ----- 398
QY 621 VTTYVPHKDTIOGAKYILAGSTKVPRAHKPRLLYNGELTCTQOSGKVNNTIYSTHGFLSN 680
DB 399 -----FKLIOAPGOI--APIILOYFG-----T 418
QY 681 LKDXGP-----DELRLAHNLMLKRPSDAMEMCRILNDEAAMNELARACLHHM----- 729
DB 419 LLDKGPLNHEHETELARPVLAON-----RIQLLEKRWYGENKLACTEALGDLV 465
QY 730 ---EVEFAIRVYRIGNV--GIWMSLEQIKGIEDYNTLAGHLAMFTNDYNLAODLYLASS 784
DB 466 KPYNTPEFALKIT--ETANVKNKVMCLSE--LGDf-----GKLATYTSQONITPD----- 511
QY 765 CPITALEMRRDIQHMDSALQIAKHLAPDOIPFISKETAIQLEFAGDYVN----- 833
DB 512 -----YVSLQONLVRYNPDQ---AAEFATOMFNSNPSINLEKIVDIFMSQ 553
QY 834 -----ALAHYKGTGDNKEHD-----EACLAGAOKMSIRMGDIRRGVNOALKHPSRY 881
DB 554 NLVQATAPALLDALKDNDNPEHSHLOTRLLEINLINAPOVA---DAIIG-NOMFTHPDRA 608
QY 882 LKRDCGAILENKOPSEAAQLEYEKL--YYDKAASV-----YIRSKNMA----- 923
DB 609 V-----IASJCERAGLVORALELDYDKPADIKRVIYHSNLLNPEWLMYFGRFSPD 658

QY 924 KYGDLLPH-----VSPKIHLOVAKAKEADGRY----- 951
DB 659 EYVDYLREMLRSLNRLQNIQVQIATRYSDLAQAOIRIEMFEKFTFEGLYYGLSIYNI 718
QY 952 ---KEAVVAYENA---KOMQSVIRIYLD--HINNPEKAAVIREQOSLDGAMVAVRFILQ 1003
DB 719 TEDPEVYKTYIOAACLIAMQFTEVERICRDNNVYNPEKVNKLKEAKLADQPLIL----- 773
QY 1004 LQDYSATIOFLVMSKCNBEAFTLAODHNMKEIYA-----DIIGSEDTTN 1047
DB 774 VCDRDVNDLVFVLYFRNNMFOF-----IELYGRINRSKTPQYVGAALDIDCDEELVQ 827
QY 1048 EDYQSTA-----LYFGEKR-----YLOAGKFFLLCGOYSBAL----- 1080
DB 828 NLMSVGVQVPYDELVVEEVERNRRLKLLPYLES---LIQSGSODRAIYDALAKIYIDSN 884
QY 1081 ---KHFLKCPSESDNVAIEMAIETVGOAKDELL-----TNQLIDHLGENDGMPKD 1128
DB 885 NNPEVFLK-----ENNFDTLIVGKYCEKRDPLYLAFLAYEKGNDTEIILNLCNENSMFKOL 940
QY 1129 AKYLFER---LYMALKQ---YREAQTAIITAREQASAGYRNAHDVLFMSYAEILKSQK 1180
DB 941 ARYLKRSDSNLMSSEVLQDSAVRRPLDQVITAVPESSDPEAVSIVYKALMEVDLPSQL 1000
QY 1181 IKI-----PSEMATNMLTILSHYILVKIHYKNGDHMKGARMKIRYANNISKFPESHIVPI 1233
DB 1001 IELLERIVLOPSSFSESNANLQNLFLTAIKADKS-----RYMEYIDLDKDYDVE 1050
QY 1234 LSTVTECHRAGLKNSAF-----SFAAMLRRPEYRSKI 1266
DB 1051 IAEIAIE---NGLYEAFRITYIKHNKHEQAMKVLVEDIYSLDRADYKAEYVQPEYMSRL 1107
QY 1267 DAKYKKKIEGAVRRPDISE 1285
DB 1108 ---AKAQLDG-IRIPDATE 1122

Search completed: May 21, 2002, 14:51:57
Job time: 137 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 14:49:41 ; Search time 16.78 Seconds

(without alignments)
3188.940 Million cell updates/sec

Title: US-09-729-653-2

Perfect score: 7257

Sequence: 1 HSLIGRCSRLGDNVAVAC.....AAQLKRIIDCTOYLRTFEEL 1382

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	182.5	2.5	1653	1 CLH_YEAST	P22137 saccharomyc
2	177	2.4	1349	1 IK13_YEAST	Q06706 saccharomyc
3	167	2.3	1332	1 IKAP_HUMAN	Q95163 homo sapien
4	161	2.2	1666	1 CLH_SCHPO	Q10161 schizosacch
5	155	2.1	1455	1 DP2L_PSEAB	P42810 pseudomonas
6	155	2.1	1455	1 DP2L_PSEAB	Q942f4 pyrococcus
7	152.5	2.1	2386	1 RAD3_SCHPO	Q02099 schizosacch
8	151	2.1	4349	1 DYHC_FUSSO	P78716 fusarium so
9	149	2.1	561	1 SR72_SCHPO	Q59787 schizosacch
10	148	2.0	1258	1 YS00_ANASP	Q847C2 anabaena sp
11	148	2.0	1557	1 DVAL_DICVT	Q24702 dictyocaulu
12	147.5	2.0	1248	1 APAF_HUMAN	Q14727 homo sapien
13	147	2.0	608	1 WDR1_XENLA	Q947f2 xenopus lae
14	143.5	2.0	609	1 WDR1_CHICK	Q93377 gallus gall
15	143.5	2.0	1681	1 CLH_CAEEL	P34574 caenorhabdi
16	142	2.0	608	1 WDR1_DROME	Q94668 drosophila
17	141.5	1.9	1675	1 CLH_RAT	P11442 rattus norv
18	140.5	1.9	1675	1 CLH1_HUMAN	Q00610 homo sapien
19	140.5	1.9	1675	1 CLH_BOVIN	P49951 bos taurus
20	140	1.9	2875	1 RRP1_TSWY1	P28976 tomato spot
21	138	1.9	567	1 YE28_METUA	Q58823 methanococc
22	138	1.9	796	1 COPP_SCHPO	Q42937 schizosacch
23	137.5	1.9	787	1 Y091_CAEEL	P41842 caenorhabdi
24	136.5	1.9	3056	1 ATP_HUMAN	Q13315 homo sapien
25	136	1.9	606	1 WDR1_HUMAN	O75083 homo sapien
26	136	1.9	1431	1 DP2L_PYRHO	Q57861 pyrococcus
27	135	1.9	1249	1 APAF_RAT	Q949v5 rattus norv
28	134.5	1.9	298	1 SNAB_BOVIN	P81126 bos taurus
29	134.5	1.9	298	1 SNAB_HUMAN	Q94115 homo sapien
30	133.5	1.8	982	1 P115_MYCPN	P46640 mycoplasma
31	133	1.8	1657	1 T0G1_HUMAN	P46640 homo sapien
32	132.5	1.8	515	1 YC2_YEAST	P25382 saccharomyc
33	132	1.8	606	1 WDR1_MOUSE	O88342 mus musculu

34	132	1.8	934	1 SYL1_SULSO	P58176 sulfolobus
35	131.5	1.8	1249	1 APAF_MOUSE	O88879 mus musculu
36	131	1.8	595	1 YB85_SCHPO	O14301 schizosacch
37	131	1.8	2547	1 FAPX_HUMAN	O93008 h probable
38	130.5	1.8	1125	1 YB62_SCHPO	O14248 schizosacch
39	129.5	1.8	566	1 YG43_YEAST	P53197 saccharomyc
40	128	1.8	1150	1 YK01_CAEEL	Q03560 caenorhabdi
41	127.5	1.8	897	1 DP01_HELPJ	Q943e9 helicobacte
42	127	1.8	1263	1 DP2L_PYRPU	P81409 pyrococcus
43	127	1.8	1597	1 CTRO_MOUSE	P49025 mus musculu
44	126.5	1.7	716	1 CA80_EUPOC	Q949x3 euploetes oc
45	126.5	1.7	3354	1 CADN_MOUSE	Q99pf4 mus musculu

ALIGNMENTS

```

RESULT 1
ID. CLH_YEAST STANDARD: PRT: 1653 AA.
AC P22137:
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Clathrin heavy chain.
GN CHC1 OR YGL206C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=91093343; PubMed=1898742;
RA Lemmon S.K., Pelliscena-Palle A., Conley K., Freund C.L.;
RT "Sequence of the clathrin heavy chain from Saccharomyces cerevisiae
RT and requirement of the COOH terminus for clathrin function.";
RL J. Cell Biol. 112:65-80(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97298309; PubMed=9153757;
RA Feuerhahn M., Simeonova L., Souciet J.-L., Potier S.;
RT "Analysis of 21.7 kb DNA sequence from the left arm of chromosome VII
RT reveals 11 open reading frames: two correspond to new genes.";
RL Yeast 13:475-477(1997).
CC -! FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYMERAL COAT OF
CC COATED PITS & VESICLES. IN YEAST, IT IS INVOLVED IN THE RETENTION
CC OF PROTEINS IN AN INTRACELLULAR MEMBRANE COMPARTMENT, PRESUMABLY
CC THE TRANS-GOLGI.
CC -! SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT.
CC -! SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC VESICLES.
CC -! DOMAIN: THE C-TERMINAL THIRD OF THE HEAVY CHAINS FORMS THE HUB OF
CC THE TRISKELION. THIS REGION CONTAINS THE TRIMERIZATION DOMAIN AND
CC THE LIGHT-CHAIN BINDING DOMAIN INVOLVED IN THE ASSEMBLY OF THE
CC CLATHRIN LATTICE.
CC -! SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL: X52900; CAA37082.1; -.
CC PIR: 272728; CAA96919.1; -.
CC PIR: A36349; A36349.
CC HSSP: P11442; 1BPO.
CC SGD: S0003174; CHC1.
CC InterPro: IPR001473; Clathrin_propel.

```

DR InterPro: IPR000547; Clathrin_repeat.
DR Pfam: PF01394; Clathrin_propel; 3.
DR Pfam: PF00637; Clathrin_repeat; 7.
DR SMART: SM00299; ClH; 7.
KM Coated pils.
FT DOMAIN 1 483 GLOBULAR TERMINAL DOMAIN.
FT DOMAIN 484 527 FLEXIBLE LINKER.
FT DOMAIN 528 1653 HEAVY CHAIN ARM.
FT DOMAIN 528 ? DISTAL SEGMENT.
FT DOMAIN ? 1653 PROXIMAL SEGMENT.
FT DOMAIN 453 469 BINDING SITE FOR THE UNCOATING ATPASE,
INVOLVED IN LATTICE DISASSEMBLY
(POTENTIAL).
FT BINDING 1219 1528 LIGHT CHAIN (BY SIMILARITY).
FT SEQUENCE 1653 AA; 187233 MW; 622F3083DF2FE315 CRC64;

Query Match 2.5%; Score 182.5; DB 1; Length 1653;
Best Local Similarity 18.4%; Pred. No. 0.0026;
Matches 288; Conservative 222; Mismatches 524; Indels 535; Gaps 75;

QY 12 LGDGNVAVCLVETGCACVLAHSSRSGSVE---MKRIFSL-----LEKTLGAPIQPAMQ 62
DB 54 LAKGNEVYRKMMGDSALHMSQWYISVRANGTYVQIFNLETKSKLSTLDEPIYF-WR 112
QY 63 KTSGNVLAVTGADYIV--KIFDRHQKRSEINLPNGCYAMDMDKDYLAVIAEKSSCIY 120
DB 113 WLSETTLGFVYARSILTSNVFD-----GN-----VNAKPOLLT 145
QY 121 LMDNTNTSOLDNGMROGMFLM-SKVGSLAVGYKGNLXINYHQTSKRTPIPLGKHT 179
DB 146 LRHALLNLT-OIINFVAKN--LDMFAVAGYLIQENGRIAGRIQLFSKO--RRTS----- 194
QY 180 KRITCGCMAENLXALGEGDKMTVSNQECPTIRQTVRSEPNXMOFELMKMDRTSAE 239
DB 195 -----QALDHVAIFTLNILEGNGSTPVQY-----FYTGNNATTTGGE 233
QY 240 SMISVY-----LGKKT--LEFLINEDPNPADEFOODFGNIYCVNYWGDRIMIG 288
DB 234 LRRIEIDHDSLPSQYKETTDFEPPDATNDFPVAQVSEKYGIIYLLTKYG----- 286
QY 289 FSCGHFVVISHTGELGDEIFQAR-----NHKDNL-----TSI 321
DB 287 -----FHLIYELTGTNLFVNRTIAESVTPAARYNHENGACINKKGOVLAVEISTSQ 339
QY 322 AVSOTLKN-----VATGCDNCKIQDLVDLKDMYIINLDEEN-----KG 361
DB 340 IVPYTLNKLNSVALALIYATRG-GLPGADDLFOKOFESLLQNDQYNAKVAASSTLRN 398
QY 362 LGTLSWTDGQ-----LLAUST--QKSLH---VFLTKPLTCD----- 395
DB 399 QNTINRLKNIQAPGALISPIILYFSTLLDKGLNKETIELAR--PVLQODRKQLEKMKL 457
QY 396 -----ACSTRIVYLSLLEVYA-----NPVEGSLPTIVSADVEPN 431
DB 458 EDKLECESEEDLDYKPPFTTLALACYLRAGAHAKYISCLAEIQEFKILPPCQKGYQPN 517
QY 432 FVAVGLYHLAVGMNRRWFYLVG-----ENA-----VKRLKDM-----EYLGTVASICLHSD 478
DB 518 FLVLE-ITSLISSPDRASEFVSLQNPETASQIDIEKIALDFEQNNIIQGTSLLDAL 576
QY 479 YAAALFEGKVOLHLIESITLDAQE-----ERETRLF 509
DB 577 KGDPDGGHLOTRVLEVLLHAPQVADAILGNIFSHYDKPTIASLSEKAGLYQALENY 636
QY 510 PAVDD--KCRILCHALTSDELI-YGTDGVVQYF-----YIEMQVNVN 549
DB 637 TDIDIKRCVVHTALPLDMLVGFYKINVSQSLACLKALMDNNIQAIVQVQVATKFS 696
QY 550 DYRHPVSVKRIFFPPNGTR-----LVFIDEKSDGVFVCPVNDATYEIPDFSTIKGV 601
DB 697 DLIGPSTLIKLFEDYNAEGLYYLASLVNLTEDKD-VVYKTI--EAAKMKQYR-EIERI 753

QY 602 LMENNPMDKGFVIAVDDKVYTYVFEHKD-TIGQAVVILIAGSTKVPAAKPLL-LYNGELT 659
DB 754 YKDN-----NHYDPERVKNFL--KDALNEDQPLVIVCDRFPFVHEMILYKSSQ-- 801
QY 660 COTQSGKVNNTIYLSHGFSLSNKQXGPDELRLPMLAHNLMLKFSQAWECRLTNEAAMN 719
DB 802 -----NKKFEITY-----VOQVNPSTQAVVALLDMD-----C--DEAFIQ 836
QY 720 ELARACLHMEV-EFAIRVYRRIGNVGYMSLEQ--IKGIED--YNLLA----- 763
DB 837 SLQSVLQGVPINELTTEVEKRNRLKILPLFEQSLQSQIQOAVYNAKAKIYDINSNP 896
QY 764 -----GH-----LAMFTNDYML--AQLLYLASS 784
DB 897 EKFLKENDQYDTLDVGHCEKRPYLAIVAEKQNDLDIRITENNSYKQYQARYL--- 953
QY 765 CPIALEMRRDQHMDSAL-----QLAKILADQIPFISKEVAIQLEFGDVNML- 835
DB 954 -----LE-RSDLDLNNKYLNDENIHRQLDSVISVGIPELDPPEVSLTYQAFMTNGLK 1007
QY 836 -----AHYERGITGDNKEHDEACLAGVAQMSIRMGDIRRGNOALKH-PSRYLKRDCGAIL 890
DB 1008 LELILELLEKIIIEPSPFENVALQGLLLS-----AIIKEPTKV-----SVYI 1050
QY 891 ENMKQF-----SEAAQLEYKGLYKKAASVYTRS--KNNAKYGDLPHVSS 934
DB 1051 EKLNDYDADELAPLCIEHDLKEAEFEIYDKHEMYGKALKVLEDDJMSIDLRASVADKINT 1110
QY 935 PKIHLQYAKKAKADG-RYKKEAVVAEYNAKQWQSVIRYLDHLNPEKAVNITREQSLDG 993
DB 1111 PELMSQIGTA-QLDGLRIPDIAEST-----IAEDPSNTEENVYIDIAE----- 1151
QY 994 AKMVARFELQLDYGSAIOFLVMS-----KCNNEAFTLAQOHKME-----I 1035
DB 1152 -----QAGKYBELLPFLMARKTLKEPRIDALILAAVELKHIHEIENLGSNVA 1202
QY 1036 YADITGSEDITNEDYQSTALYFEGEKRYLQAGKFFLLCGQYSRALKHLKCPSSSDNAI 1095
DB 1203 NLDHVGDKLFEKKEKYKARLCYSAVSNYSKLASTLYVLDYQAAADTARKA---SNTRY 1258
QY 1096 EMALIEYQAKDELTLNOL-----IDHLG--ENDGMPKDAYLFLRYALMAY 1142
DB 1259 WKLVNDACIEKKEFKLAQICGLNLIYHAEDELVEYRESNCFEELISLFPAGIGLERA 1318
QY 1143 REAAQTALITAREQSSAGNYRNAHDVLSMYAELKS-----OKIRP-----SE 1186
DB 1319 HMGFTLELAI-----LYSKYEPDKTFEHLKLFWSGRINIPKVIKRAVEQAH 1362
QY 1187 MATNIMILHSY-----LLVYKHYKNGDHMKGARMLIRAN-----NISKFP 1227
DB 1363 LMSELVFLYAHDEWDNALTLIEKSTKDLDPHAYEREVYVAVKSNDEIYKAIINFYVKFHP 1422
QY 1228 SHIVPILTS 1236
DB 1423 SLVDLILTS 1431

RESULT 2
IK13_YEAST STANDARD; PRT: 1349 AA.
AC Q06706; Q00036;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IK13 protein.
GN IK13 OR YIR384C OR L3502.7.
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97290891; PubMed=9145530;

```

RA Yajima H Tokunaga M, Nakayama-Murayama A, Hisiluma F:
RT "Characterization of IKIL and IK13 genes conferring PGKL killer
RL sensitivity on Saccharomyces cerevisiae."
RN Biosci. Biotechnol. Biochem. 61:704-709(1997).
RP (2)
RC SEQUENCE FROM N.A.
RA STRAIN=5288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favallo A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA Talch A., Trevisakis E., Vignati D., Wilcox L., Woldman P., Vaudin M.,
RA Wilson R., Waterston R.;
RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL FOR THE INTRACELLULAR KILLING PROCESS OF PGKL
CC KILLER TOXIN.
CC -1- SUBCELLULAR LOCATION: MAY BE MEMBRANE ASSOCIATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D87841; BAA20120.1; -
CC DR EMBL: U19104; AAB67278.1; -
CC DR TRANSFAC: T03453; -
CC DR SGD: S0004376; IK13.
CC KW Membrane.
CC SO SEQUENCE 1349 AA; 152989 MW; 7AEBC0BA9282F6 CRC64;

```

Dd	525	SEVDSLQVNAFINDSIYGVLLD	-----NLSTALDLDDITQDTPFLITIV	572
Qy	515	KCRILICHALTSDF---	LIYGTDTGV-----QYFYIEDW-QFYNDYRHPYSKKIRP	562
Dd	573	YDKIVY--LLRSDPFYNHLVETRGVTCQLDAEGQIMEITKPKQLVRDPF---	YKR----	623
Qy	563	DPNGRLVFEIDKSDGFVYPCVNDATYEIFDPSFTIKGVLMEMWMPDKGVFIYADDDKY		622
Dd	624	-----	VHNTSADDD-----WMSAESSELVAFG----	646
Qy	623	TYVEFHKDTIGQACVILAGST-----	KVPFAH-----KPLLY-----	654
Dd	647	--ITNNKGLFPAQVLLASVATYSLEITDSPLFLPTAQNILQFVHLNSTDRKFLPLVEEGVE		704
Qy	655	-----NGELTCQOTSGKVNNTIYLSHGLFSLMKDXG----	DELREPLMLANMLMR	701
Dd	705	DERVAIREGSGILVSVPKSSVLAQTRG--	NLETIYPRIMVLEVR---KNIMAKR	757
Qy	702	FSDAMECR-----	ILND-----EAMNELAR-----ACIHHNEVFAR	736
Dd	758	YKEAFIVCTRRINMLIDHAPFLTEMLEVEFINQIRVADYLNLFISCL--	SEDDVTKT	815
Qy	737	VYRRAGNGIYMSLEQIKGIDYVNLGLHAM-----	ETNDYNLAQD--EVLASSCP--	786
Dd	816	KYKELTYSGISKSF---	GMPRAPLTMOYIKKKMPDPTKSVNKICDAVLVNLSPD	871
Qy	787	-----IAALENRDLOHMDASLAQALAKHAPDQIPIFSKEVAI--OLEFAGD----	YYN	833
Dd	872	YKKKVLQITITAVASQNPQNLSALKILISELSE-----	EKDSGVYTLGFLDDQVNVYKS	927
Qy	834	ALAHYEKITTQDNKEHDEACIAGYAQNSIR-----	MGDIRGVNO	873
Dd	928	ALSTUDVSLA-----	LVAQKQSQMDPREYLPFLQELQDNEPLRRKFLIDIDYLGNEYK	979
Qy	874	ALKHPSRYLKRDCGALLENMKQFSEAAQIYKEYGK--	YYD--EASVYIISKMAKAVGDL	928
Dd	960	ALEHSELEIDKD--GNVSEVYDIESHDLTKHGLATRYSEKONVY--	NIYAK-----	1033
Qy	929	LPHVSSPKIHLQYKAKKADGRYKEAVAYENAKOMQSVIRIYLDHLNPEKAVNIVRET		988
Dd	1031	--HSSNQMYTDAVAAYEMGLKLEAMGAVQSAKRMEASIVAQKF--	PEEVESAUEL	108
Qy	969	OSLDGAKKVAAPFLQLDGYSAIOLFVWMSKCNNEAFPLAQOHKKMELIYDI--	IGSEDTTN	104
Dd	1087	IS-----	SLTEHHRKYVA-ADIOLELYDNVY	111
Qy	1048	EDYOSIALYFPEGKRYIQACKFPLLCQYSALKHFLKCPSSSEDNVAIEALETGQAO		110
Dd	1112	E-----AVALL-----	CAARYDIASLVAITA---KRD	113
Qy	1108	ELTTNQLDHLLEGENDGMPKD--AKYLFRLYMLAKQYREAAQTAAIITAREOSAGNYRNA		116
Dd	1137	ELL--EEVVDPLGGGFGIATLLADCKGQINSQRLRIRE-----	LRAKKEENP-----	118
Qy	1166	HDVLFSMYAEILKSQIKI--	PSEMAT	1189
Dd	1184	----YAFYGGETEQAADVVASPSETST		1206
RESULT 3				
IKAP_HUMAN STANDARD; PRT; 1332 AA.				
AC	095163;	Q9H327; Q9UG87;		
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Ikappab kinase complex-associated protein (Ikk complex-associated protein) (p150).			
GN	IKKAP OR IKAP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid:9606;			

RN [1] SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP TISSUE-Cervical carcinoma;
 RC MEDLINE=98421679; PubMed=9751059;
 RA Cohen L., Henzel W.J., Baenrele P.A.;
 RT "IKAP is a scaffold protein of the IkappaB kinase complex.";
 RL Nature 395:292-296(1998).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANT FD PRO-696.
 RC MEDLINE=21090531; PubMed=11179008;
 RA Claugenaupt S.A., Blumenfeld A., Gill S.P., Leyne M., Mull J.,
 Claugenaupt M.P., Liebert C.B., Chadwick B.P., Idelson M., Renik L.,
 Robbins C.M., Makolowska I., Brownstein M.U., Krappmann D.,
 Scheiderelt C., Maayan C., Axelrod F.B., Gusella J.F.;
 RT "Tissue-specific expression of a splicing mutation in the IKKAP gene
 causes familial dysautonomia.";
 RL Am. J. Hum. Genet. 68:598-605(2001).
 RN [3]
 RP SEQUENCE OF 961-1332 FROM N.A.
 RC TISSUE-Brain;
 RA Wandutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RL Submitted (May-1999) to the EMBL/GenBank/DBB databases.
 RN [4]
 RP VARIANT FD PRO-696, AND EFFECT ON PHOSPHORYLATION.
 RC MEDLINE=21090544; PubMed=11179021.
 RA Anderson S.L., Coll R., Daly I.W., Kichula E.A., Rork M.J.,
 Volpi S.A., Ekstein J., Rubin I.Y.;
 RT "Familial dysautonomia is caused by mutations of the IKAP gene.";
 RL Am. J. Hum. Genet. 68:753-758(2001).
 CC -1- FUNCTION: MAY ACT AS A SCAFFOLD PROTEIN THAT MAY ASSEMBLE ACTIVE
 CC IKK-MAP3K14 COMPLEXES (IKKA, IKKB AND MAP3K14/NIK).
 CC -1- SUBUNIT: INTERACTS PREFERENTIALLY WITH MAP3K14/NIK FOLLOWED BY
 CC IKK-ALPHA AND IKK-BETA.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- PPM: PHOSPHORYLATED.
 CC -1- DISEASE: DEFECTS IN IKKAP ARE THE CAUSE OF FAMILIAL DYSAUTONOMIA
 CC (FD). ALSO KNOWN AS RILEY-DAY SYNDROME OR HEREDITARY SENSORY AND
 CC AUTONOMIC NEUROPATHY TYPE III. THIS AUTOSOMAL RECESSIVE DISORDER
 CC IS DUE TO THE POOR DEVELOPMENT AND SURVIVAL, AND PROGRESSIVE
 CC DEGENERATION OF THE SENSORY, SYMPATHETIC AND PARASYMPATHETIC
 CC NERVOUS. FD INDIVIDUALS ARE AFFECTED WITH A VARIETY OF SYMPTOMS
 CC SUCH AS DECREASED SENSITIVITY TO PAIN AND TEMPERATURE,
 CC CARDIOVASCULAR INSTABILITY, RECURRENT PNEUMONIAS, VOMITING CRISES,
 CC AND GASTROINTESTINAL DYSFUNCTION. IT IS PRIMARILY CONFINED TO
 CC INDIVIDUALS OF ASHKENAZI JEWISH DESCENT, WITH AN INCIDENCE OF
 CC 1/3600 LIVE BIRTHS.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 1286.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF044195; AAC64258.1; -;
 DR EMBL: AF153419; AAC43369.1; -;
 DR EMBL: AL049945; CAB43219.1; ALT_FRAME.
 DR MIM: 603722; -;
 DR MIM: 223900; -;
 KW Phosphorylation; Disease mutation.
 FT VARIANT 696 696 R -> P (IN FD; MILD PHENOTYPE;
 FT PHOSPHORYLATION IS REDUCED).
 FT /FTID=VAR_011327.
 FT CONFLICT 304 304 R -> W (IN REF. 2).
 FT CONFLICT 312 312 K -> E (IN REF. 2).
 FT CONFLICT 754 754 P -> L (IN REF. 2).
 FT CONFLICT 961 961 C -> G (IN REF. 3).
 FT CONFLICT 1072 1072 S -> C (IN REF. 3).
 FT CONFLICT 1158 1158 P -> L (IN REF. 3).
 FT CONFLICT 1320 1320 I -> V (IN REF. 3).

SQ SEQUENCE 1332 AA; 150191 MW; 3FD65FAB554DA923 CRC64;
 Query Match 2.3%; Score 167; DB 1; Length 1332;
 Best Local Similarity 18.5%; Pred. No. 0.018;
 Matches 250; Conservative 188; Mismatches 456; Indels 456; Gaps 65;
 OY 51 TWLGAPGFAMQKTSNGYLA-----TGADYVIFDRHG--QKRSEINLPQNCVANDW 102
 DB 199 FWRG-----DGFQFAVSVCPEFGARK-VWVWNEFALQSTSE-PVAGLPALAW 246
 OY 103 DKDGYLAIVAKKSSCTILMDANTKTQSLDNGMDQKSFILMSKVSFLAVGTGKGLX 162
 DB 247 KPSGSLISTQDKP-----NQDDYFERKNG-----L 273
 OY 163 IYNHOTSRIPIYLGKTRITCGCNAAE-NLXALGGEKMTVSNQEDDTIRQTVRSEP 221
 DB 274 LHGHT---LPEFL-KDEYKVNDLIMNASSVLALED-----LQREKSSITKTVQVLT 324
 OY 222 XNMFELMKMDRTSAESMISVILGKTLFPLNEDPNADLEFQDPFGNIVCYNMWG 281
 DB 325 VGNHYMYLK-----QSLSFSTCGSKIVSL-KMDPVTYRLHVLQCGWHYLAIDWHM 375
 OY 282 DGRIMIGFSCGHFYVISTHTGE-----LGEIIFQARHK 315
 DB 376 TTDRSGVGNSSDLSNVAVIDGNRVLVTVPROTVPPMCTYOLLEPHPVNOYTFIAHQR 435
 OY 316 DNLTSIAVSQTLNKVAT--CGDNCIKIODLV-----DLKDMVYLMD 356
 DB 436 SN--DLAVLDASNQISYVKCGD-CPSADPTVKGAVGSGFVCLRTPLHLEKRYIQFEN 492
 OY 357 EENKG-----IGTSLWTDGGLLALS---TQSGLHFLTKPLPLGDACSTRIAYTSL 407
 DB 493 NEDQVNPDLKGLTLTWIEEDVFLAVSHSEFSRPSYIH-HLT-----AASSEM----- 538
 OY 408 LEVTYANVEGELPTVSDVDEPNVAVGLYHLAAGMNNRAFYVLGENAVKKLDMETL 467
 DB 539 -----DEBGLNVSSSAVD----- 554
 OY 468 GTVASICLHS--DYAALFEKGVLHLESEILDAQERETRLP----- 510
 DB 555 GVITSLCCNSKTKSVVLQADQIKFYLMESPSLAIKPKMNSGPRVPRPCTQTELAM 614
 OY 511 -----AVDKCR-----ILCHALTSPLIYGTDTGVQVFIYEDNQF----- 547
 DB 615 IGEBCVLGLDRCRPFINDIEVASNITSFAYVDEFLTTHSHRCQCFCLDASFKTLQ 674
 OY 548 ----VNDYRHPYSVKIRPPDNGRLVFIDEKSDGFVYCPVNDATVYEIIPDFSPTIKGYLM 603
 DB 675 AGLSSNHYSHGEVLKV---ERGSRTIVV-VPODKVILQMRGLVYVHHNALVLAOI- 729
 OY 604 ENWPMKDGVFAYDDDKVYTYVFKHDTIOGAKVILLAGSTKYFAHKLPLLYNGELTCQOQ 663
 DB 730 RKM-LDKLIMF-----KEAFECMRKLININPI-YDHNPKVFLGNVETFIKQ 773
 OY 664 SKCVNNIYLTSHGFLSNLKDQCPDELRLP---LAHNLMLKRFSDAMEMCRILNDEAMN 719
 DB 774 IDSVNHNIL---FETELKE--EDVTKTMYPAPVSSVYLSHDPGKNIDLVCDAMRAVM 827
 OY 720 ELA---RACL---HHM-----EVEFAL-RVYRIRGN-----VGIYMSLEQIGIDY 759
 DB 828 ESINPHKVCISLITSHVAKTTPLELIVLOKYNHELOGNAPSDPDVAVSABEALKYLLHLVDY 887
 OY 760 NLLACHLAFNDYNYLAODLYLASSCPAALERMRLDQHMDSALQALANHLADQIPFTSK 819
 DB 888 NELYDH-SLGYDYF---DLVL-----MVAEKSQKDPREYLPFLNTLKKMETNYQRFITD 937
 OY 820 EYALDLERAGDYVNAHLAYEKGITGDKN---EHDACLAGVAGQMSIRMGDIRRGVQNALK 876
 DB 938 KY-----LKYERKAIGHLSKCGPEYFPECNLTKDKNL-----YNEALK 976
 OY 877 --HPSRVILKRCQ---GATLENNKOFSEAAOLYERGLVYDKAASVYIRSKNNAKVGDLDP 930

```

Db 977 LVPSSQOYQDISIAYGEHLMQEHMEYEPAGLMPARCAHEKALSAFLTCGN----- 1028
Oy 931 HVSPKIHQYAKAKADGRYKA--VVAENAKOMOSYIRYLHLNPEKAVNIREF 988
Db 1029 -----KQALCVAAQNLFTKDQVGL-----GRTLAGKIVEOR 1060
Oy 989 OSLDGAKVNAFFLQDGYSAIOFLVSKCNNEAFTLAQHNKMEI----- 1035
Db 1061 KHIIDAAVLEE---SAQDEEAVLLLEGAAEALRLVYKKNRDLIETNVKPSILEAQ 1117
Oy 1036 --YADITSEDT-----NEDVQSIALYFEGEKRYIQAGKFFL----- 1071
Db 1118 KNVMAFLDSQATFSRHKRLVRELKQAOQAGL--DDEVPHQGESDLFSETSSVVG 1175
Oy 1072 --LGGVYSA-----LKHFLKCPSSBDNVAIEMAIETVGO-----AKDE 1108
Db 1176 SMSKSKYHSNRISARSSKNRKAERKHSLEKESPLEDLALLELSVONTENLKE 1235
Oy 1109 LITNQLIDHLGENDGMPKADAYLFRLYMALQYREAAO---TAITIR----- 1154
Db 1236 -----VYHIL-----KVLFLPEFDEGRELQAFEDTLQLMERSLPEITWTLTY 1278
Oy 1155 EEOSAGNTRNADVLFSMAELKSOKIKIP 1184
Db 1279 CONSATPVLGPNSTANSIMASYOQKTSVP 1308

```

RESULT 4

CLH_SCHPO STANDARD: PRT: 1666 AA.

```

AC Q10161;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Probable clathrin heavy chain.
GN SPMC26A3.05.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_Taxid=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Mclean J., Harris D., Bartell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC COATED PITS AND VESICLES (BY SIMILARITY).
CC -1- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC VESICLES (BY SIMILARITY).
CC -1- DOMAIN: THE C-TERMINAL THIRD OF THE HEAVY CHAINS FORMS THE HUB OF
CC THE TRISKELION. THIS REGION CONTAINS THE TRIMERIZATION DOMAIN AND
CC THE LIGHT-CHAIN BINDING DOMAIN INVOLVED IN THE ASSEMBLY OF THE
CC CLATHRIN LATTICE.
CC -1- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 269240; CAA93228.1; .
DR HSSP: P11442; 1BPO.
DR InterPro: IPR001473; Clathrin_Propel.
DR InterPro: IPR000547; Clathrin_Repeat.
DR Pfam: PF01394; Clathrin_Propel; 5.
DR Pfam: PF00637; Clathrin_Repeat; 7.

```

DR SMART: SM00299; CLH: 7.
 KN Hypothetical protein: Coated pits.
 SQ SEQUENCE 1666 AA; 190017 MW; 16A6B84FEA2A7EA9 CRC64;

Query Match 2.2%; Score 161; DB 1; Length 1666;
 Best Local Similarity 18.7%; Pred. No. 0.059;
 Matches 251; Conservative 167; Mismatches 445; Indels 476; Gaps 65;

```

Oy 192 LVALGEGDKMITVSNQEGTITQTVRSEPHXNM-QEFLMKMDRISAASMSI-----VVL 246
Db 15 LASVGIQPSFSGFANVTLESDEKYVCVRDNPNGVNVQVYLDLDPSSVNLRRPISADSVILH 74
Oy 247 GKTLFELNLBNPDNPADLEFOODFC-----NIYVYMYGGGRIMGF-----SCGFVY 296
Db 75 PKKRIIALAQAQOLQVFDLEAKAKINSYMNQDVYMTWISDS--YIGAVTDTSVRHMTV 132
Oy 297 -----ISTHTGELGOEI--FOARHKDNLTSIAVSQTLNKVATGDNCI-----KI- 340
Db 133 SGSDPVKMFDRHSLSLNGTQIISYKSNYNEWFTLIGISSRDNRIA--GNLQYSKRKVS 190
Oy 341 QDLVLDKDMYVL---NLDEBNKGIGTSLWTDGQLLALSTO--KSLHVFILKPLIG- 394
Db 191 QPLESHASAFAYIQEGVDHE-----VOYLALASRLPTGS-----KLSIVEY 232
Oy 395 DACSTRIAVLTSLLEVTVANPEGELPIYVSVDVEPNFAV-----GLYHL-----AVG 443
Db 233 DRPNPNPAPATITVDLFFPEPEAVNDPPIAIEIGSTYNVAVYTKGFIHYVDETAKCIY 292
Oy 444 MN---NRAMFYVIGENAVKRLKDMEYLGTVASICLHSDYAALFEKGVOLHLEISEILDA 500
Db 293 MNRVSGESIFVTTHAKSVNGLAINRKQGVLSINPE-----TIIPYLSNLND- 342
Oy 501 QEERETRLPAPVDDCRILCHALTSDFLYGDTGVYQFYIEDMOPFVNDYRIPSVYKXI 560
Db 343 -----PGI--AVRAMSHA-----NLPGADNLYMQGF--QQLAAGNYSBAKVAAS 384
Oy 561 PDDPNGRTRLVFIDEKSDGVYCPVNDATYEIPDSPTIGVLNEMPMKGVFIAYDDK 620
Db 385 SP-----KGLRTSVIDQ----- 398
Oy 621 VYTVYFHKDTIOGAKYIAGSTKVPRAHKPLLLYNGELTCOTGSKVNNIYSTHGLSN 680
Db 399 -----FKLIQAAPGDI--APILQYFG-----T 418
Oy 681 LKDGP-----DELRLAHNLMLKRSDAEMKRIILDEAANWELARACLIHM----- 729
Db 419 LLDKGLNEHETIELARPVLAQN-----RIQULEKMYGENTLACTEALGDLV 465
Oy 730 ---EVEFAIRVYRIGNV--GIYMSLEQIKGIEDYMLAGHLAMPNDYNLAQDYLASS 784
Db 466 KRYNPFALKTY-ETANVNRKVMCLSE--LQDF--GKLATYYSOONITPD----- 511
Oy 785 CPITALEMRRDQLQHWDSALQALAKHLAPDQIPFSKEAYALQLEFAGDVN----- 833
Db 512 -----YVSLILQNLVRYNPDQ---AAPFAQGMFNSNSINLEKLVDFMSQ 553
Oy 834 -----ALATYKGTIDDKNEHD-----EACLAGYAQMSIRMGDIRRGVNAKLKPSRY 881
Db 554 NLVQAQATLFLDALKDNDEHSHLQTRLEINLINAPOVA---DAILG-NQMFTHDRA 608
Oy 882 LKRDCAILENMKQFSEAQLYEKGL-YVDKASV-----YIRSKMNA----- 923
Db 609 V-----IASLCEKRGVLQVRALELYDKPADIRVIVHSHLNPMLMTFSFSPD 658
Oy 924 KVGDLPLH-----VSSPKIHQYAKAKADGRY----- 951
Db 659 EYVDYLREMLRSNLRONQIIVQO IATRYSDLVQAQRIIEFEKFKFEGLYYVLGSIANI 718
Oy 952 ---KEAVVAEANA---KOMOSYIRIYLD-HLNPPEKAVNIYREIOTSDGAKVNAFFLQ 1003
Db 719 TEDPEVYVRYQIAACIMQNFETVEERICRDNVNYNPEKRVNLKEARLADQLPLIL----- 773

```


RA Hellig R.;
 RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome
 structure and evolution";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: POSSESSES TWO ACTIVITIES: A DNA SYNTHESIS (POLYMERASE)
 CC AND AN EXONUCLEOTIC ACTIVITY THAT DEGRADES SINGLE STRANDED DNA
 CC IN THE 3' TO 5' DIRECTION. HAS A TEMPLATE-PRIMER PREFERENCE WHICH
 CC IS CHARACTERISTIC OF A REPLICATIVE DNA POLYMERASE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate -> N diphosphate
 CC + (DNA)(N).
 CC -1- CATALYTIC ACTIVITY: DEGRADATION OF SINGLE-STRANDED DNA. IT ACTS
 CC PROGRESSIVELY IN A 3' TO 5' DIRECTION, RELEASING 5'-
 CC PHOSPHOMONONUCLEOTIDES.
 CC -1- SUBUNIT: HETERODIMER OF A LARGE SUBUNIT AND A SMALL SUBUNIT (BY
 CC SIMILARITY).
 CC -1- PTH: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
 CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
 CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE ARCHAEAL DNA POLYMERASE II FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AJ248283; CAB49044.1; -
 CC InterPro: IPR003586; HIntc.
 CC InterPro: IPR003587; HIntn.
 CC InterPro: IPR002203; Intein.
 CC SMART: SM00305; HIntc: 1.
 CC SMART: SM00306; HIntn: 1.
 CC DR PROSITE: PS00881; PROTEIN_SPLICING; FALSE_NEG.
 CC KW Transferase: DNA-directed DNA polymerase; DNA replication; Hydrolase;
 CC Nuclease; Exonuclease; DNA-binding; Multifunctional enzyme;
 CC Autocatalytic cleavage; Protein splicing; Complete proteome.
 CC FT CHAIN 1 954 DNA POLYMERASE II LARGE SUBUNIT, 1ST PART
 CC (POTENTIAL).
 CC FT CHAIN 955 1139 PAB POLC INTEIN (POTENTIAL).
 CC FT CHAIN 1140 1455 DNA POLYMERASE II LARGE SUBUNIT, 2ND PART
 CC (POTENTIAL).
 CC SO SEQUENCE 1455 AA; 165709 MW; BB092E7BB2FE0D21 CRC64;

Query Match 2.1%; Score 155; DB 1; Length 1455;
 Best Local Similarity 21.0%; Pred. No. 0.11;
 Matches 161; Conservative 90; Mismatches 288; Indels 226; Gaps 35;

QY 732 EFARIYRRIGANGVYMSLEQIKGIEDYNL-----LAGHL 766
 DB 100 EQAVRTALALTEGIVSA--PIEGIANVKIKRTNWDNSEYALVYAGPIRSSGGTAQAL 157
 QY 767 AMFTNDY---NLADLYLASSCPITALENMRDIOHMDSAQLAKHLAPDOIPISEKVAI 823
 DB 158 SYLVGVYVRKRLGLDRFKPSEKHEEMVEVDLYH-RAVTRLOHYHSPREVRAMKRIRI 216
 QY 824 QLEFAGDYNALAHAYKGTGDNKEHDEACLAGVAQMSIRMGDI---RCGVNALKHPSP 880
 DB 217 EI-----TGEATDDEVEVSHRDVPGVETNQLGAILVLAEGVLQAKKLVK 262
 QY 881 VKRDCGAILENMKOPSEAAQLYE-----KGLYVYKASVYI 917
 DB 263 YIDKMGIEGHEWLTKEVEAKEKPEKKEESLAESTLEETVEYDMGYT---SLYQ 318
 QY 918 R-----SKMAK--VGDLPLHVSSPK---IHLQYAKAK----- 945
 DB 319 KPEETIAPSKVAKYVIGG-RPLFSDPSKRGFRRLRGSRASGATWGINPATMILVDE 377
 QY 946 -----EADGRYKAVAVENAKQMO-----SVIRITLDLHNPEKAVNIYRET 988
 DB 378 FLAIGTOLKTERPG--KGAVVTPTTIEGPIVRLKDGSVLRV--DDYNLALKVREDVEEI 433

QY 989 QSLDGAKMVARFLOGLDYGSAIOLYVMSKNNEAFILAQOHKMEIYADIIGSEDTTNE 1048
 DB 434 LYTGDA-----VIARGDEVENNQOTLLPANYCEEMWLEEVKALKETYE--VHLEPTEN 485
 QY 1049 DYOSIALYFEGEKERYLOAGKFFLLCGQYSRALKHFLLKCPSSSDNVAIEMAI---ETVGOA 1105
 DB 486 EESI-----EASDYLEIDPEF-----LKEMLRDP-LRVKPPVELAIHSEVILGIP 531
 QY 1106 KDELLTNQIDILLGENDGMPKDAKYLFRLYMAKQYREAAQTAAIIIAEEOQSNAGNTR 1165
 DB 532 LHPYTT-----LYNVSVPKDVEXKIMRL--LKNYAEI-----EWSNF 566
 QY 1166 HDVLFSGVARELSOKIRIPESEMTNMLHSYLLVYKHHVNG-----DHMGARMLIRVAN 1221
 DB 567 RGIRFAKKIYISOEKLGDGSKRTLELGLPHT-----VADGVYDYPPAALALPPLGN 619
 QY 1222 NISKPSHIVPILTSYVIECHRAGLKNARSF--AAMLRPEYRSKIDAK----- 1269
 DB 620 LMNEFMAKPLIYATIDINENNEIKLRDGIWIGARGREKAKERKMKPPVGLPPIGL 679
 QY 1270 ---YKKIEGMVRRPDISIEEATTPCFCKFLPPECCLCPGCKNSIPYCIATGRHMK 1326
 DB 680 AGGSSSRDIKKAAEGKVAEVEIAFFCKPCGCHVP--EHLCPNC-----GTRKELL- 728
 QY 1327 DDMTCVPHC--DFPALYSELKIMLNTESPQPCMSERLNAOQLKXI 1369
 DB 729 --W-VCPRCNAEYPESOAE-----GIYVTCPCNVKVLRYPAKKRI 765

RESULT 7
 RAD3 SCHPO STANDARD; PRT; 2386 AA.
 ID RAD3 SCHPO Q92059; Q92059; Q92391;
 AC 020259; Q92059; Q92391;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA repair protein rad3.
 GN RAD3 OR SPBC216.05.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=972;
 RX MEDLINE=97133293; PubMed=8978690;
 RA Bentley N.J., Holtzman D.A., Flagg G., Keegan K.S., Demaggio A.,
 RA Ford J.C., Hoekstra M., Carr A.M.;
 RT "The Schizosaccharomyces pombe rad3 checkpoint gene.";
 RL EMBO J. 13:6641-6651(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Lyne M., Rajandream M.A., Barrell B.G., Oliver K., Harris D.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 711-1781 FROM N.A.
 RX MEDLINE=93012978; PubMed=1398093;
 RA Seaton B.L., Yucel J., Sunnerhagen P., Subramani S.;
 RT Isolation and characterization of the Schizosaccharomyces pombe rad3
 RT gene, involved in the DNA damage and DNA synthesis checkpoints.";
 RL gene.119:83-89(1992).
 CC -1- FUNCTION: INVOLVED IN G2 ARREST FOLLOWING DNA DAMAGE WHERE IT
 CC PHOSPHORYLATES CHK1. IT IS ALSO INVOLVED IN THE DEPENDENCE OF
 CC MITOSIS ON THE COMPLETION OF DNA REPLICATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Y09076: CAA70297.1: -
DR EMBL: U76307: AAC49607.1: -
DR EMBL: AL049556: CAB40165.1: -
DR EMBL: X63544: CAA45106.1: -
DR PIR: S25834: S25834.
DR InterPro: IPR003151: FATC.
DR InterPro: IPR003152: FATC.
DR InterPro: IPR000403: p13_p14_kinase.
DR InterPro: IPR001440: TPR.
DR Pfam: PF02259: FAT; 2.
DR Pfam: PF02260: FATC; 1.
DR Pfam: PF00454: p13_p14_kinase; 1.
DR SMART: SM00146: PI3K; 1.
DR PROSITE: PS00915: p13_4_KINASE_1; FALSE_NEG.
DR PROSITE: PS00916: p13_4_KINASE_2; FALSE_NEG.
DR PROSITE: PS0290: p13_4_KINASE_3; 1.
KW Transferrase; kinase; DNA damage; DNA synthesis; DNA repair;
KW Nuclear protein.
FT DOMAIN 2018 2326 PI3K/PI4K.
FT CONFLICT 198 198 T -> S (IN REF. 1).
FT CONFLICT 526 526 L -> P (IN REF. 1).
FT CONFLICT 1779 1781 SEC -> LRM (IN REF. 3).
SQ SEQUENCE 2386 AA; 273526 MW; DDEA343FEA3872D3 CRC64;

Query Match 2.1%; Score 152.5; DB 1; Length 2386;

Best Local Similarity 17.3%; Pred. No. 0.34; Mismatches 563; Indels 533; Gaps 78;
Matches 281; Conservative 244;

QY 25 GCA-----CVLHSSRSGSSEVEMKRIPLSLEKT--WLGAPIQ-----FAMQKTSGNV- 68
DB 510 GCEVNSFCFLDERSLFPIRYHELFCALKNPDISSSVKQSLLDGFRHSQHCSNN 569
QY 69 -----LAVTG-----ADIVKIFDRHGQKRSEINLPQNCVAMDMDKQ-- 106
DB 570 KESMLSLREFIMKALASTSRCLRVAAKVLPIF-----IKGPNNDIYEFHESKA 620
QY 107 ---DVLAVIAEKSSC-----IYLMQANTNKTOLDNGMDQKSFLLMSKVGSFLAVGTAK 158
DB 621 LIENLKLIAVENTALILEVILISLSRVVE-----EELHFVILEYSSVINSGLTY 674
QY 159 GNLATYVNHQ---TSRRIPVLGKHTKRITCGCWAENALAGGEDKMITVSN-----Q 207
DB 675 QGIGLSALQOQIASTRHSIYW---QLSPYWPVTSVAIVQGMGKKPNIASLFAQLMNTS 729
QY 208 EDDTIRQVQVREPRXNMOPFLMKMDR---TSAESMISVVLGKTLPLFLNLE----- 258
DB 730 EDDLIRIQVATYLP---FLVLTNKKALIVRIALSOSDV---ATLCLTNNKKILASLL 781
QY 259 -PDNPADLEFQODFGNIVCYNMGGRIMIGFSCGHF-----VYSTGTGELGEIFQA 311
DB 782 TTDHP-----NLESVMILLSTATSDFEKVDLTSLRSPISITVELLOL 826
QY 312 RNHKNLTSIAVSQTLNKVATGDCNCIKIQLDVLKDMYVILNDEENKGLTSLWTDDG 371
DB 827 --YQNDVHEKIEALRKVAMIVSQVNDDELSNKKELVDFN---NHILGILAEFSN- 879
QY 372 QLLALSTGRSLHV-----FLTKLPILGDACSTRATVLTSLLEVTVAN----- 414
DB 880 ---TJLNDKGKTSINEKIKTIIVGIEKMLSLCGAAVKLGIPQILSNLSQAFQENHLRFYAI 936
QY 415 -----PVEGELPTIVSDVDPNPNVAVGLVHL 440
DB 937 KAWSLIILATEPEYSSTAGSLVLLPLFYLPEQAEALVYQIR-----DTISSDTHRC 991
QY 441 AVGNM-----NRAMFYVLGENAVKLLDMETLGTVASI--CLHSDYAAALFEKQVOLH 491
DB 992 LOGLKWALPTSLDSACFSLKAKEIFCSQJNEDFYSELQSIICLNENPNPVYGLQKLE 1051

QY 492 L-----ISEILLDAQEERETRLFPVAVDCKRILCHALTSDFLIY----- 530
DB 1052 LFFQAKVDELHPTLNDINSNEVLD-----QLRLCLDCC--VKYASTNMQSYLAANK 1102
QY 531 ---GT-----DTGVVOYFY--IEDMOVVNDYRHPVSKKIFPPDNKRIYLF 571
DB 1103 LSELGAIIDPSRAKAOHIIKETVYVLNDFENGESLSKILDFMQ---SOLIP---AFLYT 1154
QY 572 IDEKSDGFVYCPVNDATYEIPDF-----SPTI---KG--VLWMW-----PMDKCVFTA 615
DB 1155 TDTKAGFL-----AYALQEFILKIGFKSAVINKKKLLVTYTEHMSLPLDSKRVLLP 1207
QY 616 YDDDKVY-----TYVFKKDTI-----QGA-----KVILA 639
DB 1208 FLTSKYHLPIPIKIDIRYPIYKENVTIHTWMQLFSKLMEVAHSAQAEKIFGICSKYVD 1267
QY 640 GSTKVPFAKRPILLYNGLTCTQTSQKVNNTYLTSTGFLSNLKDQRPDLRPLAHN---- 696
DB 1268 QEVNIPCFLLPFLVNLVLT--ESELVNVK--IEEFQVLVNIQPPDGLNSVGQQRVYS 1322
QY 697 ---LMLKRFSDAEMCRILNDEAAMNELARACLHMEVEFAIRYRRIGVNGVMSIE 751
DB 1323 FVDVFEKIVDYLNKMLRMKKKNMWDERSAIARKENYNSVEDATSSSSISKVESFLSRF 1382
QY 752 QIKGIEDVNLG-----HL-----AMFTNDYNLAQDLYLASSCP--IALE 791
DB 1383 PSKTLGIYSLNCGFNARALFYWEQHIRNATAPYAALESYRVLOETIYAGIDDPDEIEAVS 1442
QY 792 M-----RRDLOH-----WDSAL---OLAKHILAPDQIFLSKEVAI---QLEFAGY 831
DB 1443 LNFHDYSFDQQLIDHNSGTWALSCEYELTIQKDE---NKKRAKILSLMSLDQSGHY 1497
QY 832 VALAHYEKIGITGDNKEHDEACIAGV--AQMISIRMGDIRGVNOALKHSRYLKRDCGAI 889
DB 1498 ESUVLSDSFLIINDHREYSKMLNLGIEASWRSLSIDSLKCLS---KSNLESFEKAGSTI 1554
QY 890 LENMKQFSAADLYEK--GLYDKAASVYIRSKNAKV-----DLLPHVSEKITHL 939
DB 1555 FYQYLKDSFAELTERLOPLVDAATAI---ANGAHSAYDCDYTL---SKLHA 1602
QY 940 ---QYAKAKAAD-----GRYKEAV-----VAYEAKOMOSYIRTY 971
DB 1603 INDFSRATIDTDIVSDNIDIVLRRLISOVAPRGKRRHQHLSHLVGYEKFEMTKTAELY 1662
QY 972 LD-----HLNPD--EKAVINIVRETOSLDGAKV---ARFQLQGDYGSALQFLVMSKCNNE 1022
DB 1663 LEIARISRKNGQFORAFNAILKAMDLDKPLATIEHAQMMHOGQRKATSEINFN- LNN 1721
QY 1023 AFTLAQOH-----NKMELY-----ADIISEPTJNDYOSIALYFEG 1059
DB 1722 MFDLVDEHEERPKNRKRETLGNPLKGVFLKTLKWLKQAQQLDKLDLEYHNAVEIYSEC 1781
QY 1060 ERRYLQAGKFFLLCGOYSRALKFLKCPSEEDNVAIEMALIEYVGOAKDLELTNQLDHL 1119
DB 1782 ENTHYLLGHRRRLMYEERKL-----PVNQSEKFLSGELVTRTI 1821
QY 1120 GE-----NDGMPKDAKYLFRLYMAL-----KQYREAQTAIIAR 1154
DB 1822 NEFGRLSYLYGTNHIYESMPK---LTLWLDFGAELRLSKDGEKYFREL-----IIS 1872
QY 1155 EEOASGNRNMADVLFESYMAELKSQKIKIPSEMAINMLLHSHYILVKIHVKG-----DH 1209
DB 1873 RRRKS-----DELNMSNVCRLSMKIPDYFF--LVALSOMISRVCJPNKRYKILFH 1920
QY 1210 MKGARMILRVANNISKFPSHIYPIILTSFY--ICHRAGAKNSAFSFAAMLRPREYSKIDA 1268
DB 1921 I-----IANVASYFOETLMQMATIKTSQKRSLRGKISLVNLVHSRKLMSKSDVI 1972
QY 1269 K 1269
DB 1973 K 1973


```

RESULT      8
DHYC_FUSSO  STANDARD:      PRT: 4349 AA.
AC   p78716:
DT   01-NOV-1997 (rel. 35, Created)
DT   01-NOV-1997 (rel. 35, Last sequence update)
DT   30-MAY-2000 (rel. 39, Last annotation update)
DE   Dynein heavy chain, cytosolic (DHYC).
GN   DHCL.
OS   Fusarium solani (subsp. pist.) (Nectria haematococca).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Nectriaceae; Haematonectria.
ON   NCBI_TaxID=109625;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F213;
RA   Inoue S., Aist J.R., Turgeon B.G., Yoder O.C.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC   -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC   MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC   ORGANELLES ALONG MICROTUBULES.
CC   -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC   INTERMEDIATE AND LIGHT CHAINS.
CC   -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
CC   -----
DR   EMBL: U84215; AAC33176.1; -.
DR   HSSP: P03069; 12IU.
DR   InterPro: IPR004273; Dynein_heavy.
DR   InterPro: IPR001482; GSPIL_E.
DR   InterPro: IPR002017; Spectrin.
DR   Pfam: PF03028; Dynein_heavy; 1.
DR   Pfam: PF00437; GSPIL_E; 1.
KW   Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT   DOMAIN            459      480      COILED COIL (POTENTIAL).
FT   DOMAIN            1178     1215     COILED COIL (POTENTIAL).
FT   DOMAIN            1266     1293     COILED COIL (POTENTIAL).
FT   DOMAIN            1334     1354     COILED COIL (POTENTIAL).
FT   DOMAIN            1560     1577     COILED COIL (POTENTIAL).
FT   DOMAIN            1640     1670     COILED COIL (POTENTIAL).
FT   DOMAIN            2048     2076     MICROTUBULE-BINDING (POTENTIAL).
FT   DOMAIN            2194     2217     COILED COIL (POTENTIAL).
FT   DOMAIN            3186     3294     COILED COIL (POTENTIAL).
FT   DOMAIN            3420     3477     COILED COIL (POTENTIAL).
FT   DOMAIN            3774     3807     COILED COIL (POTENTIAL).
FT   DOMAIN            1946     1953     ATP (POTENTIAL).
FT   NP_BIND            1946     2239     ATP (POTENTIAL).
FT   NP_BIND            2604     2611     ATP (POTENTIAL).
FT   NP_BIND            2946     2953     ATP (POTENTIAL).
SO   SEQUENCE            4349 AA; 493453 MW; FCB3C7152B36A1BF CRC64;

Query Match      2.1%; Score 151; DB 1; Length 4349;
Best Local Similarity 18.2%; Pred. No. 1;
Matches 234; Conservative 172; Mismatches 432; Indels 446; Gaps 61;

```

```

QY 146 SKVGSFLAVGTAKGNLXIYNHOTSRIKIPVGLKHTKRTTCGMNAENLALGDEKMTIVS 205
DB 613 SKENALFVPRKIRGAIOEQONO-----LMHVKKO-----AINGLHEFKQO 653
QY 206 NOEGDTIROTVRSEP--XNMOFFLKKMDRTSAESMISVYLKKTFFLLNLEPDNPA 263
DB 654 YGSHETNAQAQLRDLDPVSGAIIWQAQIEFOLDGVRKKEAVLGPDMWHTTECHKLOEBS 713
QY 264 DLEFODDFGNVYCN-WYGD-GRIMTGFSGHHVYSTHTGELGQIFQARHKMDLTSI 321
DB 714 EL-EKOKLDTARIYEAMIDVGRKISTIS-----GQ-LFE----- 746
QY 322 AVSQTLNKVATCGDNCKIKIDVLDKMYVILNLDENKGLGTSWTDGOLLALSTORG 381
DB 747 -----IARVRSAG-----GILELYVNPDPVYITLFKEFR 775
QY 382 SLHVFLLKPLIIGDACSTRAYLTSLLEVTVANPVGELPIYVSVDYERPNFVAVGLYHLA 441
DB 776 NL-----TWOSYSVPHAVTVTSKDAKRVYPVAVSLMESVRTLSQTLRQIS 820
QY 442 VGMNNNAFVYVGENAVKLLKMEYIGTVASICLHSDYAAALFEGVYOLH---LISEIL 498
DB 821 -----VMGEE-----SVLLFGYRNDVYKLISEG--VPLRWESFINSHEL 857
QY 499 DAQEEERTR-LFPAVDCKRILCHALTSDFLIYGTDTGVQVYVIDMQVFNDYRHPVSY 557
DB 858 FTSDNROTFRPLRG-----GTPGLAKNTESHKGFIRKFAAHSV 898
QY 558 KIIEPDNCTRLVEIDEK--SDGEVYCPVNDATYEIIPDSPTIKGLWMENWPMKGVFLA 615
DB 899 -----LOQKAVSLNFIHATVEQALKEL-----WTCPEEAAFNHS 932
QY 616 YDDDKYVTVFHKDTIQGA--KYTLAGSTKVPRAHPRLLYNGLTCQOSQGVNNIYLS 673
DB 933 -----RLDTTQAADVQNLNLEQVYNLDF-----WNGLSKVQSILTLRLQSA 974
QY 674 THGFLSLNKDXGP-DELRPLAHN-----LMLKRFSD 704
DB 975 VHAMIAFEDDDPDDDMRRKRVNNNNNEAKPDGPTMKRYVELAMRNOVIYLDLLEFARA 1034
QY 705 AWEM-----CRI-----LNDEANMELARACLHMEVEFAIRVYRR 740
DB 1035 SWELHLHEWLGIVCNLRKIKATRYQMSLTFTTANDEPRFDLPSECAGLLQ-----RVY-- 1087
QY 741 IGVVGVMSLEQKIGIEDYNLAGHLAMPTNDYNL-AQDLIYLAASSCPYLALEMRRLQHW 799
DB 1088 ---VSEKKRLHEVSAYVDKWL-----QFOSLMDLQSEQVYDA-----LGEQLPRW 1129
QY 800 DSAIQAKHLAPDQIFISKEVAYIQLEFAGDYVNALAH-----YEKGITGDNKEHD-----E 851
DB 1130 ---LQLLQETIKRSTFTDQVDS-----RAGCHLTIDYDQVQTVYNAKYDQWQHE 1176
QY 852 ACLAGVAOMSIRMGDIRRGVNAQLKHPSRVLRKDCGAILLENMKQFSEAQLYEKGLYDVK 911
DB 1177 ILMKFASRLGRNRRREINAEIEKARKH-----LES--QSSDASTAQAQVQITV 1222
QY 912 AASVYIRSKNMKAVGOLLPHVSSPKIHLQYAKAKKEADGRKKEVAVYENAKQMSVIRYI 971
DB 1223 VQSCRRNVKTWAPETIDFPGOGSTLVROKQFPN--DWLHTEOI-----DSQWALEKETI 1275
QY 972 LDHLNRPKAVNIVRETOSLDGAKNVARFELQGDYSAIQFLVMSKCNNEAFT-LAQOH 1030
DB 1276 -----EKKSRIYQVQDTALQAKIYAE-----DKLINERIAELAAQM 1311
QY 1031 NKMEIYADIIGSDTINEDYQIAL-YFEGEKRYLOAGKFFLLCGOYSRALKHFLLKCPSS 1089
DB 1312 NEEKPVS-----GTIQPVASATLSFSRSRISKLODD-----AQMVAKAKKALDIPAS 1359
QY 1090 EDNVALEMAIEIV-----GQAKDELLT-----NQLIDHLLGENDGM 1125
DB 1360 PD-TSLEATILEEYRDQSVWSNLSITWASLNTERDVLMTAVOPRKIRSKYDDILKSTKEN 1418
QY 1126 PKDAKYLFRLYMALKO-----YREAQTAIIIAREEOSAGNYRNAHDVLFVSMVAELSKQRI 1181

```

```

Db 1419 PS-----RMQYAFEFHVQILGFLKVNSTLSDKSDAIRERHMKI-----YKQIKPQKR 1470
QY 1182 KIPSEMAT-----NLMITHSYLVK 1201
Db 1471 FSPSSMTLGDVMDLNVATEVTVK 1494

RESULT 9
SR72_SCHPO STANDARD: PRT: 561 AA.
AC 059787:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Signal recognition particle 72 kDa protein homolog (SRP72).
GN SPC320.10.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Weller H., Wambutt R.;
RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE
CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPASMIC
CC RETICULUM MEMBRANE. SRP72 BINDS THE 7S RNA ONLY IN PRESENCE OF
CC SRP68. THIS RIBONUCLEOPROTEIN COMPLEX MIGHT INTERACT DIRECTLY WITH
CC THE DOCKING PROTEIN IN THE ER MEMBRANE AND POSSIBLY PARTICIPATE
CC IN THE ELONGATION ARREST FUNCTION (BY SIMILARITY).
CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE
CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,
CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE SRP72 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL022245: CAAL8312.1: -.
DR InterPro: IPR001440: TPR.
KM Signal recognition particle; Ribonucleoprotein.
SO SEQUENCE 561 AA: 63281 MW: 635A3954/CA51BDB CRC64:

Query Match 2.1%; Score 149; DB 1: Length 561;
Best Local Similarity 20.6%; Pred No. 0.067;
Matches 119; Conservative 87; Mismatches 151; Indels 220; Gaps 31;

```

```

QY 1056 YFEGEKRYIQAGKFFLLGQYRALKHFLKCPSSD-----NVAIMEAIE----- 1100
Db 192 FE-----AYLQA-----GDNEKASKESLKI--SRKCDJETSQALFVNNLISMSIDNPIYI 238
QY 1101 -----TVGAKDELLTNO-----LIDHLIGENDGMPKPAK----- 1130
Db 239 SFRDLHGNTLEKALSSSLASQKKQFIRNLALDDMAKGKORSYRKEKKRNPERSITYFTIL 298
QY 1131 -----YFRLYMAKQYREAQTAIIAREQDSAGYRNADHLEFSMYA 1174
Db 299 LREETKSLISPKKLPGYLENF--KSDSDNIVALLMQHHSINGNFRGA---LSIY- 350
QY 1175 ELKSQKIKIPSEMATNMLTHSYIIV-----KTHVKNQDMKARMLIRYANNISKFPSP-- 1228
Db 351 ---QKLRISLASQSLSYLSPGLVGLDGLAHYKIQSTGFSPQSLLHEAXANWRKQOQSC 406
QY 1229 -----HI--VPIIT-----STVICHRAGLKNASAFSFA 1255
Db 407 AKLLICTNSLNAHDERAVSTIQDDMSYIDLLQMKGFISLIVSCKYAAAL----- 457
QY 1256 MLMRPEYRSKIDAKYKKIKEGWVRPPDISIEEATTP 1292
Db 458 CYLDKEIESGMD-KYLVPTKDLITGIDVDDIETRGVP 493

RESULT 10
Y500_ANASP STANDARD: PRT: 1258 AA.
AC 08YVC2:
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein alr2800.
GN ALR2800.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21595285; PubMed=11759840;
RA Kaneo T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Matenabe A., Iriduchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishibe Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -1- SIMILARITY: CONTAINS 15 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP003590: BAB74499.1: -.
DR PROSITE: PS00678; WD_REPEATS_1: 9.
DR PROSITE: PS50082; WD_REPEATS_2: 14.
DR PROSITE: PS50294; WD_REPEATS_REGION: 1.
KM Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 55 93 WD 1.
FT REPEAT 640 679 WD 2.
FT REPEAT 682 721 WD 3.
FT REPEAT 724 763 WD 4.
FT REPEAT 766 807 WD 5.
FT REPEAT 809 849 WD 6.
FT REPEAT 850 889 WD 7.
FT REPEAT 892 931 WD 8.
FT REPEAT 934 975 WD 9.
FT REPEAT 976 1017 WD 10.

```

FT REPEAT 1019 1059 MD 11.
 FT REPEAT 1060 1101 MD 12.
 FT REPEAT 1103 1143 MD 13.
 FT REPEAT 1144 1183 MD 14.
 FT REPEAT 1186 1227 MD 15.
 SQ SEQUENCE 1258 AA; 139513 MW; 45DF03B91170C451 CRC64;

Query Match 2.0%; Score 148; DB 1; Length 1258;
 Best Local Similarity 19.3%; Pred. No. 0.25; Indels 178; Gaps 29;
 Matches 119; Conservative 89; Mismatches 232;

23 ETGCAVLHSSRGSSVEKRIEFLSEKTLWLGAPIOFAMOKTSGNYLATGADYIVKIPD 82
 DB 661 DTDCHVRWVEVASKKLLICRHS-----NW-----VAFVFSFDGELLACGADENKWLMS 712
 QY 83 RHGQRSEINLPNCV-----AMDWKDGDVLAIVAEKSSCIYLMADNTKTSQ 131
 DB 713 VR-----DGVCIKTLTGHEHEVEFSAVAFHPDGETLA-SASGDKTKILMDIODGTCLQ 762
 QY 132 LONGRHDSFILMSKVSFLAVGVKGNLXVYHOTSRIKIPVLGKHTKRITCGCMNAE- 190
 DB 763 TLTGHTDWRCVAFSPDQCTLLASADHTTKLMDVSQGCRLTKLSHTGWMRSVAFSADG 822
 QY 191 NLXALGCEDKMTVSNQE-----GDTIRQTVRESEPMXNQFLMKMDRTSA---A 238
 DB 823 QTLASSGGRTIKIMNYHNGECLTKYIGHTNSVYSAISP-DKILVSGSGDRTIKLMDC 881
 QY 239 ESMISVILGKTLFLNLNEPNPADLEFQDFGNIVCN-----WY 280
 DB 882 QNHICL---KTLH-----GHTNEVCSVAFSPDQTLACVSLDSQVRLMNCRTGQCLKAWY 933
 QY 281 G---DGRIMGFCGHHVYVSTHTG-----ELGGEIQRNHNKNLSIANV---SQT 327
 DB 934 GMTDALPVAFSPDQCTLLASGNDKTVKLMWQTKYISSLEGHTDFYIGAFSPDSQTL 993
 QY 328 NKVATCGDNCIKI-----QDLVLDKD-MYVIL----- 353
 DB 994 AASAT--DSSVRLMNSTGOCFOILLEHTDWYAVVFNHOGKIATGSDCYKLMKNISF 1051
 QY 354 ----NLDENKGLGTLTMDGOLLALSTQRCSLHVFPLKPLTLDACSTRIAYL---- 404
 DB 1052 GCKLKLSEHSDKILGMAMSPQQLLASASADSVR-----LMDCTGRCVGLIRGH 1103
 QY 405 TSLLEVTYANPVEGELPIYVSVDVEPNFVAVGLYHLAVGMNNRMYVIGENAVKTKLMD 464
 DB 1104 SNRVSAISP-NGELIATCSTD-----QTVKIMWQOG---KCKLTL 1142
 QY 465 E-YLGTVASICLHSDYAALFEKGVOLHLESEILDQAEERETRLPAVDKCRILC--- 520
 DB 1143 TGHNTMNVFPIAFSP-----GKI-----LASASHDQIVRIMDVMTGKCHHICIGH 1187
 QY 521 -NALTSDFLIYGTGCV 537
 DB 1188 THLYSS--VAFSPDGEV 1203
 RESULT 11
 DVAL_DICVI STANDARD; PRT; 1557 AA.
 AC 024702;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DVA-1 polypeptide precursor (Antigen-1) (Allergen-1) (NPA).
 GN DVA-1.
 OS Dictyocaulus viviparus (Bovine Lungworm).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 OC Trichostrongyloidea; Dictyocaulidae; Dictyocaulinae; Dictyocaulus.
 OX NCBI_TaxID:29172;
 RX MEDLINE=96089829; PubMed=8538702;

RA Britton C., Moore J., Gilleard J.S., Kennedy M.W.;
 RT "Extensive diversity in repeat unit sequences of the cDNA encoding
 RT the polypeptide antigen/allergen from the bovine lungworm
 RT Dictyocaulus viviparus.";
 RL Mol. Biochem. Parasitol. 72:77-88(1995).
 RN [2]

RP FATTY ACID/RETINOID-BINDING.
 RX MEDLINE=95370256; PubMed=7642601;
 RA Kennedy M.W., Britton C., Price N.C., Kelly S.M., Cooper A.;
 RT "The DVA-1 polypeptide of the parasitic nematode Dictyocaulus
 RT viviparus. A small helix-rich lipid-binding protein.";
 RL J. Biol. Chem. 270:19277-19281(1995).
 CC -1- FUNCTION: HAS HIGH BINDING AFFINITY FOR FATTY ACIDS AND RETINOLIDS.
 CC -1- SIMILARITY: TO A.SUNM ABA-1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: 002568; AAC47403.1;
 KW Signal; Polypeptide; Lipid-binding; Retinol-binding; Repeat;
 KW Glycoprotein; Antigen; Allergen.
 FT CHAIN 1 21
 FT PEPTIDE 22 1557
 FT PEPTIDE 61 213
 FT PEPTIDE 214 334
 FT PEPTIDE 335 480
 FT PEPTIDE 481 606
 FT PEPTIDE 607 737
 FT PEPTIDE 738 869
 FT PEPTIDE 870 1003
 FT PEPTIDE 1004 1074
 FT PEPTIDE 1075 1208
 FT PEPTIDE 1209 1342
 FT PEPTIDE 1343 1475
 FT PEPTIDE 1476 1545
 FT PEPTIDE 1546 1557
 FT CARBOHYD 997 997
 FT VARIANT 946 946
 FT VARIANT 948 948
 FT VARIANT 955 955
 FT VARIANT 974 974
 FT VARIANT 990 990
 FT VARIANT 1543 1543
 SQ SEQUENCE 1557 AA; 182600 MW; 33717911CF55BDAB CRC64;

Query Match 2.0%; Score 148; DB 1; Length 1557;
 Best Local Similarity 18.6%; Pred. No. 0.35;
 Matches 256; Conservative 193; Mismatches 502; Indels 424; Gaps 71;

31 HSSRSQSSVEMKRIEFLSEK--TWLGAPIOFAMOKTSGNYLATGGA---DYIVKIPDRHG 85
 DB 408 HHHHRHLAVRRRHLVAIEFLDMLKPE-----QKHLEKLENSGAHFDVIAEVKKFYG 462
 QY 86 ----QKRSEINLPNCVAMDMD-----GDVLAIVAE----- 114
 DB 463 LPEEKIELKAKFKSCQCYDVAKVATSEEMNDIMKHESKNHSDLMKRLTELENRLTED 522
 QY 115 -----KSSCIYLMDA-NTNK--TSQDNGRHDQSFILMSKVSFLAVGVKGNLX 163
 DB 523 QKHTIEHREVCCGLMEVQNTNKHOSLSEADAVLSMTDEDEKVKAIYETSNROT 562
 QY 164 YNH-----QTSRKIPVLGKHTKRITCGCM-AENLXALGSD-KMTVSNQEGDTIRQTV 217
 DB 583 YDELTKMESSEDEVRKAKATEKLEAACKHGTNI--LGEENVDLIRMKNGATFEISN 640
 QY 218 RSEPMXNQFLMK---MDRTSAESMISVILGKTLFLNLNEPNPADLEFQ--DFGN 273

```

Db 641 RVDE-----LLEGITDSRKEKAYRMSKLC--KKIYSLGHSK-----OLQOYDEEN 684
QY 274 IV--CYNW-----YDGR1-----MIGFSGHFVY 297
Db 685 VLOKLTWLDSDQKNELTMSDNKEIKYKIIDYDGTIGEVKEKAEVLOLACHNHYIK- 743
QY 298 STHGELCOEFPQARNKNDLTSA--VSQILNKVATGDCNCKIOD---LVLDKMWY 351
Db 744 SIVEEKMEIKQJKECKSSSEIKAKVEDVYINQIS--DESIRSRADALLVCRIKIGI 800
QY 352 ILNDEENKGGT-----LSWTDGOLLASTQSGSLNHLFLKPLIGDAGCTRIAY 403
Db 801 VKRLRDNSEHSLSEAMERLTLWLSDOCKYIKS-----IDYDNKVLXY 846
QY 404 LLSLEVTANPVEGELPITYSVDEPNFVAVGLYHNAVGNMNRAMFVYLGENAVKIKLD 463
Db 847 -EKIME-----FDDAIGETK-----QKAKKELKD 870
QY 464 MEYLGTVASICLHSDYAAALF---EGKVQHLISEIL-DAQEETRLFPVAVDKCR-- 517
Db 871 -----ACKH--YVVDLIGEENGMLREKENGASNEALATKVEIMEIATDETRKA 919
QY 518 ILCHALTSDPLIGTGVGYOYFYEDMOFVNDYRHPVSKKIFPDNGTRLVFIDEKSD 577
Db 920 QAMASTSCRKVY---GVVQ-----RFRDHHENHLEBALF---KHFTWLNESOK 964
QY 578 GFVYCPVNDATYELPDSPTIKGLWMENMPDKGVFIAYDDKVYTYV-----FHKDT 630
Db 965 SOL-----KTIYESEDRBLHKV-WEFFEGAGLRASNASKKIYGVAKRFRDHHENH 1018
QY 631 IQGAKVILAGSTKVPFAKPLLYNGELTCOTQSGKVANNITLSTGFLSNLKDQXPDEL 690
Db 1019 LDEA-----LEKYLTLWNEE---QKSOMKTY-----ESSDRE-- 1048
QY 691 PMLAHNLMLKRFSPAMKCRILNDEAMNELARACLIHM-----EVEFAIRVYRIGNV 744
Db 1049 -----ALYKVLFEFEATGEVKEKAVALKSACRHYIKDYGKREKKEIKEMESG-- 1100
QY 745 GIVMSLEOI-KGIEDYNLAGHLANFTNDYLAODLYLASSCPT--AALEMRDLQWMD 800
Db 1101 ---VSTEIISRKVDEF-----IAMITDEKKAKALIRASSACKIYGVAKRFRDHHENH 1151
QY 801 SALQ-LAKHLAPDOIPETISKEVATOLEF--AGD-----YVAAALHYEKGTGDNKEHDEA 852
Db 1152 NLEBPLEKTYLT-----WLNESQKSOMKTIYSGDREALKVYLEFFE-AAEGEVEKAAV 1205
QY 853 CIAGVAQMSIR--MGDIRGVNOALKHPRSVYLKRDGAILNEMKOFSEAAOLYELKLYD 910
Db 1206 ELKSACRHYINDYIGDK-----AAKIKEMKESGVSTEE----- 1239
QY 911 KAASVYIRSKMMAKVGDLPHVSSPKHLOQAKAKEADGRYKEAVVAYENAKQOWSVIRI 970
Db 1240 -----ISK--KYDEFIAMITDE--KKAKALIRASNACK--IYVAKR---LRR 1279
QY 971 YLDHLNDEKRVNIVRETOSIDGAKMVARFLQJGDYGSATQFLVMSKNNEAFTLAQO 1030
Db 1280 DHHHEHNEEMAGKYLSMMSDQOQVKKI-----YGTGR-----LATYN 1320
QY 1031 NKMEIYADIISEEDTJNEDYOSIALYFEGEKRYLOAGKFFLLCGOYRALKHFLKCPSE 1090
Db 1321 KVMLEFESVPEDEK-----EK-----ATSOLKACRHYITDFIGK 1355
QY 1091 DNVAIEMAIETVGAQKDELTLNOLIDHLLGENDGMPKDAKYLFLYMLAKOYREAAQTAI 1150
Db 1356 DNLAIVIKEMESG--ATNEAIGEKIDEPFAGLDDEOK-----AQOARASACKKI 1404
QY 1151 IIAEEDSAGYRNAHNVLFMY-----AELKSQIKITPSMATNMLHSHYILYK 1201
Db 1405 YGVASRRRREHYEIDVEALSKYLTWLNESQAKELIKQJKEK--DEKQF---IGKRIMEF 1458
QY 1202 IHVKNQDHMKGARMILIRVANNISKFPSSHIVPILITSTVECHRRAGLKNASFPAAM 1256
Db 1459 FELTSGDCKEKARQOLKAA-----CKHYVKNYGEKEAALKLIKLDKSGISLEEM 1507

```

```

RESULT 12
ID APAF_HUMAN STANDARD. PRT: 1248 AA.
AC 014727; Q9UBZ5; Q9U5B8; Q9U560; Q9U561; Q9U562; Q9U563;
AC Q9U564; Q9U565; Q9U566; Q9U567; Q9BZ6; Q9UGN8; Q9UGN9;
AC Q9UGP0; Q9UNC9;
DT 15-JUL-1998 (Rel. 36, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Apoptotic protease activating factor 1 (Apaf-1).
GN APAF1 OR KIA0413.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3), AND PARTIAL SEQUENCE.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=97410306; PubMed=9267021;
RA Zou H., Henzel W.J., Liu X., Lutschg A., Wang X.;
RT "Apaf-1, a human protein homologous to C. elegans CED-4, participates
RT in cytochrome c-dependent activation of caspase-3."
RT Cell 90:405-413(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 4 AND 5).
RC TISSUE=Peripheral blood, Heart, and Cervical carcinoma;
RX MEDLINE=99373149; PubMed=10441496;
RA Hahn C., Hirsch B., Jahne D., Duerkop H., Stein H.;
RT "Three new types of Apaf-1 in mammalian cells."
RT Biochem. Biophys. Res. Commun. 261:746-749(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=T-cell;
RX MEDLINE=99292765; PubMed=10364241;
RA Saleh A., Srinivasula S.M., Acharya S., Fishel R., Alnemri E.S.;
RT "Cytochrome c and dATP-mediated oligomerization of Apaf-1 is a
RT prerequisite for procaspase-9 activation."
RT J. Biol. Chem. 274:17941-17945(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Kidney;
RA Benedict M.A., Nunez G.;
RT Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 385-1248 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=98116655; PubMed=9455477;
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VIII.
RT 78 new cDNA clones from brain which code for large proteins in
RT vitro."
RT DNA Res. 4:307-313(1997).
RN [6]
RP SEQUENCE OF 810-864 AND 866-883 FROM N.A.
RA Roberts D.L., Dalgleish R., Cohen G.M., MacFarlane M.;
RT "The mammalian CED4 homologue, APAF1, exists as two distinct forms in
RT human cells."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 1-138 FROM N.A. (ISOFORM 1/4/5).
RA Won H., Lee J.-W., Ohr H.-H., Kim D.-U., Chung K.-S., Lee M.,
RA Yoo H.-S.;
RT "Cloning of variant Apaf1."
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [8]
RP APAF-1-MEDIATED OLIGOMERIZATION.
RX MEDLINE=98315378; PubMed=9651578;
RA Srinivasula S.M., Ahmad M., Fernandes-Alnemri T., Alnemri E.S.;
RT "Autoactivation of procaspase-9 by Apaf-1-mediated oligomerization."
RL Mol. Cell 1:949-957(1998).

```

[9] INDUCTION BY E2F AND p53.
RA MEDLINE-21283226; PubMed-11389439;
RA Moroni M.C., Hickman E.S., Denchi E.L., Caprara G., Collì E.,
RA Ceccconi F., Mueller H., Helin K.;
RA "Apaf-1 is a transcriptional target for E2F and p53.";
RA Nat. Cell Biol. 3:552-558(2001).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (1.3 ANGSTROMS) OF 1-97.
RX MEDLINE-20013059; PubMed-10543941;
RX Vaughn D.E., Rodriguez J., Lasehnik Y., Joshua-Tor L.;
RX "Crystal structure of Apaf-1 caspase recruitment domain: an alpha-
RX helical Greek key fold for apoptotic signaling.";
RL J. Mol. Biol. 293:439-447(1999).
RN [11]
RP STRUCTURE BY NMR OF 1-97.
RX MEDLINE-20047184; PubMed-10578182;
RX Day C.L., Dupont C., Lackmann M., Vaux D.L., Hinds M.G.;
RX "Solution structure and mutagenesis of the caspase recruitment domain
RX (CARD) from Apaf-1.";
RL Cell Death Differ. 6:1125-1132(1999).
CC -1- FUNCTION: Oligomeric Apaf-1 mediates the cytochrome c-dependent
CC autocatalytic activation of pro-caspase 9 (Apaf-3), leading to the
CC activation of caspase-3 and apoptosis. This activation requires
CC ATP.
CC -1- SUBUNIT: Monomer. Oligomerizes upon binding of cytochrome c and
CC dATP. Oligomeric Apaf-1 and pro-caspase-9 bind to each other via
CC their respective NH2-terminal CARD domains and consecutively
CC mature caspase-9 is released from the complex.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- ALTERNATIVE PRODUCTS: 5 isoforms; 1/Apaf-1XL (shown here), 2/Apaf-
CC 1L, 3/Apaf-1S, 4/Apaf-1M and 5/Apaf-1XS; are produced by
CC alternative splicing.
CC -1- TISSUE SPECIFICITY: Ubiquitous. Highest levels of expression in
CC adult spleen and peripheral blood leukocytes, and in fetal brain,
CC kidney and lung. Isoform 1 is expressed in heart, kidney and
CC liver.
CC -1- INDUCTION: By E2F and p53 in apoptotic neurons.
CC -1- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -1- SIMILARITY: CONTAINS 13 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: CONTAINS 1 NB-ARC DOMAIN.
CC -1- CAUTION: Ref.7 sequence differs from that shown due to a
CC frameshift in position 109.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF013263; AAC51678.1; -
DR EMBL: AJ243003; CAB55579.1; -
DR EMBL: AJ243004; CAB55580.1; -
DR EMBL: AJ243005; CAB55581.1; -
DR EMBL: AJ243006; CAB55582.1; -
DR EMBL: AJ243007; CAB55583.1; -
DR EMBL: AJ243008; CAB55584.1; -
DR EMBL: AJ243009; CAB55585.1; -
DR EMBL: AJ243010; CAB55586.1; -
DR EMBL: AJ243011; CAB55587.1; -
DR EMBL: AJ243012; CAB55588.1; -
DR EMBL: AJ243107; CAB56462.1; -
DR EMBL: AF134397; AAD38344.1; -
DR EMBL: AF149794; AAD34016.1; -
DR EMBL: AB007873; BAA24843.1; -
DR EMBL: AJ133643; CAB55085.1; -
DR EMBL: AJ133644; CAB55086.1; -
DR EMBL: AJ133645; CAB55087.1; -
DR EMBL: AF248734; AAK28401.1; ALT_FRAME.
DR PDB: 1CY5; 01-DEC-99.
DR PDB: 1CWM; 21-JAN-00.

DR	MIM: 602233; -;	CARD.
DR	InterPro: IPR001315;	CARD.
DR	InterPro: IPR002182;	NB-ARC.
DR	InterPro: IPR001680;	WD40.
DR	Pfam: PF00619;	CARD. 1.
DR	Pfam: PF00931;	NB-ARC; 1.
DR	Pfam: PF00400;	WD40; 11.
DR	PRINTS; PR00320;	GPROTEINRPT.
DR	SMART; SM00320;	WD40; 10.
DR	PROSITE; PS50209;	CARD; 1.
DR	PROSITE; PS00678;	WD_REPEATS_1; 4.
DR	PROSITE; PS50082;	WD_REPEATS_2; 9.
DR	PROSITE; PS50294;	WD_REPEATS_REGION; 1.
KW	Apoptosis; ATP-binding; Repeat;	WD repeat; Alternative splicing; 3D-structure.
FT	DOMAIN	1 90 CARD.
FT	NP_BIND	104 415 NB-ARC.
FT	REPEAT	154 161 ATP (POTENTIAL).
FT	REPEAT	613 652 WD 1.
FT	REPEAT	655 694 WD 2.
FT	REPEAT	697 738 WD 3.
FT	REPEAT	741 780 WD 4.
FT	REPEAT	796 836 WD 5.
FT	REPEAT	838 877 WD 6.
FT	REPEAT	880 919 WD 7.
FT	REPEAT	959 998 WD 8.
FT	REPEAT	1001 1040 WD 9.
FT	REPEAT	1042 1080 WD 10.
FT	REPEAT	1083 1122 WD 11.
FT	REPEAT	1125 1164 WD 12.
FT	REPEAT	1175 1212 WD 13.
FT	DNAI1	95 98 POLY-SER.
FT	VARSPLIC	99 109 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT	VARSPLIC	575 575 E -> ETGFGESKR (IN ISOFORM 5).
FT	VARSPLIC	824 866 MISSING (IN ISOFORM 3, ISOFORM 4 AND ISOFORM 5).
FT	VARSPLIC	1113 1154 MISSING (IN ISOFORM 5).
FT	CONFLICT	108 108 S -> P (IN REF. 7).
FT	CONFLICT	134 134 N -> S (IN REF. 7).
FT	CONFLICT	145 145 G -> C (IN REF. 2; CAB55587).
FT	CONFLICT	161 161 S -> F (IN REF. 2; CAB55586).
FT	CONFLICT	370 370 Y -> T (IN REF. 2; CAB55581).
FT	CONFLICT	383 383 Y -> H (IN REF. 2; CAB55586).
FT	CONFLICT	544 544 F -> T (IN REF. 2; CAB55584).
FT	CONFLICT	580 580 A -> T (IN REF. 2; CAB55580).
FT	CONFLICT	608 608 R -> C (IN REF. 2; CAB55585).
FT	CONFLICT	620 620 H -> R (IN REF. 2; CAB55583).
FT	CONFLICT	639 639 L -> F (IN REF. 2; CAB55579).
FT	CONFLICT	708 708 T -> A (IN REF. 2; CAB55584).
FT	CONFLICT	742 742 H -> R (IN REF. 2; CAB55586).
FT	CONFLICT	746 746 V -> A (IN REF. 2; CAB55586).
FT	CONFLICT	757 757 L -> P (IN REF. 2; CAB55462).
FT	CONFLICT	795 795 E -> G (IN REF. 2; CAB55581).
FT	CONFLICT	798 798 E -> G (IN REF. 2; CAB55587).
FT	CONFLICT	825 825 D -> A (IN REF. 2; CAB55587).
FT	CONFLICT	871 871 S -> L (IN REF. 2; CAB55581).
FT	CONFLICT	876 876 A -> T (IN REF. 2; CAB55581).
FT	CONFLICT	949 949 I -> V (IN REF. 2; CAB55585).
Query Match	2.0%; Score 147.5; DB 1; Length 1248;	
Best Local Similarity	19.7%; Pred. No. 0.27;	
Matches 116; conservative 85; Mismatches 210; Indels 179; Gaps 28;		
Oy	102 WDKGDVLAIVAEKSSC-----ITYMDANTNTKSOLDNGMRDMSFLMKSVCGLAVGT	156
Db	623 FESEDCGRIR-----SCGADKTQVFVKATGEKLEIKRAHDEVLACCAFSPTDRIEAFCS	676
Oy	157 VSGNLXIVYHOTSRKIPVLGKHTKRTGCC--WNLEN-LXALGED--KMITVSNOE	208
Db	677 VKKKAKINSMTGELVHYTHDEHSOYNC-CHTNSSHLLLATGSSDFCLKMLDLNOKEC	735
Oy	209 -----GDTLRQTGVSEPNMOF-----LKKMDRTSAESMTISVYLGGKTL---FF	253

```

Db 736 RNTMFGHTNSVNHCRFSPDDKILLASCADGTLKLDATSANF-----RKSINVKQFF 787
QY 234 LNLNEDPNADLEFOODEGNIY-CYNWYGDG-RIMIGFSCGHFVYIHTHGLGGEI--- 308
Db 768 LNLNEDP-----QEDMEVYKCCSMASDARIWA-AKNHIFLFDIHTSLGLGEIHG 838
QY 309 -----FOARHKNKDNLTSAVSQTLNKVAATCGDNCKIKIODLVLDKMDVYILNDEEN 359
Db 839 HNSTIQYCDFPSQNH---LAVVALSQ-----YCYE-----LMTWDSRS 873
QY 360 KGL---GTLNW-----TDDGOLLASTORGSLHVFILFPLTLDGACSTRIVATLSLEY 410
Db 874 KVADCRGHLNWGHVGFSPDGSSFTLSDQTLRLMETK---KYCKNSAVMLKQEVYD 928
QY 411 T-----VANVEGELPTVSDVEPEFPAVAGVHLAVGNRRAMFEVY 452
Db 929 VQGENEVWVLANVDIRRLQDLINGRTGQIDYLEAQVSCCCLSPHLOYIAFGENGCA---- 984
QY 453 LGENAVKRLKDMEVYIGTASICLHSDYAAALFEKVQVQLHLESLIDAOERETRLFPAY 512
Db 985 -----TEILELVNRRIFQSRFQHKKTVMHIOFTADEKTLISSSDDAEIQVMNQ 1033
QY 513 DDKC-----RLCHALTSDFLIVGTDTGVQVYFIEDMOFVNDVRHPSVKK 559
Db 1034 LDKCIFLNGHOETVYKDFRLKNSRLLSWSEDT-----VKYWNITTKMK-----EK 1079
QY 560 IFPDNGTRLVEFIDEKSDGFVYCPVNDATYEIPEFSPITIKVLENNWMD 609
Db 1080 DFVCHQGVTL-----SC---DISHDTRKFSSTADTAIKMSRD 1115

```

```

RESULT 13
ID MDRL_XENLA STANDARD: PRT: 608 AA.
AC 09W/F2:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE WD-repeat protein 1 (Actin interacting protein 1) (XAIP1).
GN MDRL OR AIP1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
NC NCB1_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99230343; PubMed=10212149;
RA Okada K., Obinata T., Abe H.;
RT "XAIP1: a xenopus homologue of yeast actin interacting protein 1
RT (AIP1), which induces disassembly of actin filaments cooperatively
RT with ADF/cofilin family proteins."
RL J. Cell Sci. 112:1553-1565(1999).
CC -1- FUNCTION: INDUCES DISASSEMBLY OF ACTIN FILAMENTS IN CONJUNCTION
CC WITH ADF/COFILIN FAMILY PROTEINS.
CC -1- SIMILARITY: CONTAINS 11 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE AIP1 FAMILY OF WD-REPEAT PROTEINS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)

```

```

CC EMBL: AF124140; AAD22062.1;
DR InterPro: IPR001680; WD40.
DR Pfam: PF004400; WD40. 9.
DR PRINTS: PRO0320; GPROTEINRPT.
DR SMART: SM00320; WD40. 11.
DR PROSITE: PS00676; WD_REPEATS_1; 1.
DR PROSITE: PS50082; WD_REPEATS_2; 4.

```

```

DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Actin-binding; Cytoskeleton; Repeat; WD repeat.
FT REPEAT 56 95 WD 1.
FT REPEAT 100 143 WD 2.
FT REPEAT 145 184 WD 3.
FT REPEAT 187 226 WD 4.
FT REPEAT 232 271 WD 5.
FT REPEAT 318 359 WD 6.
FT REPEAT 363 401 WD 7.
FT REPEAT 436 482 WD 8.
FT REPEAT 487 526 WD 9.
FT REPEAT 530 569 WD 10.
FT REPEAT 575 607 WD 11.
SQ SEQUENCE 608 AA; 66092 MW; C8C0F8A474A759F0 CRC64;

```

```

Query Match 2.0%; Score 147; DB 1; Length 608;
Best Local Similarity 19.1%; Freq. No. 0.1;
Matches 112; Conservative 81; Mismatches 212; Indels 182; Gaps 24;

```

```

QY 40 EKRIIFSLLEKTWLGAPIQFAMQKTSQNYLAVTGADYIK-----IFDRH----- 84
Db 4 ELKKYFASLPQMERGVAKTLIADGPKGNFLYINGKSVIIRINDNPALDIYTEHHAQAVY 63
QY 85 -----GKRSEINL-----PGNCVAMDWDKGDVLAIVA-- 113
Db 64 ARVAPSGFIASGDIPTSLIMDTTQKEHLKYEYQPFAGKIKDIATWEDSKRIAVVGE 123
QY 114 -EKSSCIYIMDANTN-----KTSQDLN----- 134
Db 124 RKRFASVFLMDGGSSVGEITGNIKVINSVDIKOTRPRYRLVTSDDNCAFLGPPFKFE 183
QY 135 GMRDMSF---LLMSKVSFLAVGTGKGLXLYNHQTSRKIPVLG---KHTKRITCGGN 188
Db 184 TMSDSRFVNCVRFSPDGSKLASAGADGQIFLYDGKTGKVCVSLGSAHGDGIYAVGMS 243
QY 189 AENLXALGE-DKMITVSNQSGDTIRQT-QVRSEPKNQVFLMKDDRTSAAESMISVYL 246
Db 244 PDGTOLLSASGCTTKIKMDVANSVAFTFNLSVDLDQDLGCLMOKD-----YLLSVSL 297
QY 247 GKTLFFPLNLPDNPADLEFQDGFNGIVCY---NWYGGRIIMFSGGHFVISTHGE 303
Db 298 -SGYINIDKKNPAPRLVY-IKGHKKSIOCMVTVNNSDGRSTIYTSHDGHIWYDAETGE 355
QY 304 LGQELFOARNHKNDLTSAVSQTLNKVATCG-DNCIKIOTDYL----- 344
Db 356 --NDFFTKGHTNQVSSMDL-DGCGQLITCSMDITLRTSLSKSYSSSESVMQVQKC 412
QY 345 -----DLKDMYVILNDEE-----NKGIGTIS-WTDDGOLLAL 376
Db 413 VAVSGGVVTVYICQIYLLKDKKKVFADISLDYEPKVAALHKGSTVAVAGDAGKVDLY 472
QY 377 SGQRSLHVFILKPLILDAGACSTRIVATLSLEYVAVNVEGELPTVSDVEPFAVAG 436
Db 473 STQGSLSLDEGKTLIPAKCAVYDLDVSHDCAFLAVDAKV-----VT 514
QY 437 LYHLAVGNRRAMFEVYLGENA---YKRLKDMYELGT-----VASICL 475
Db 515 VFSVADGYSEKNSYV--GHAAKALSVAMSPDNEHFASSGMGMWGCL 559

```

```

RESULT 14
ID MDRL_CHICK STANDARD: PRT: 609 AA.
AC 093277:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE WD-repeat protein 1 (Actin interacting protein 1).
GN MDRL
OS Gallus gallus (chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

```

```

OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99156857; PubMed=10036186;
RA Adler H.J., Wlinski R.S., Gong T.-W.L., Lomax M.I.;
RT "A gene upregulated in the acoustically damaged chick basilar papilla
  encodes a novel WD40 repeat protein."
RL Genomics 56:59-69(1999).
CC -1- FUNCTION: INDICES DISASSEMBLY OF ACTIN FILAMENTS IN CONJUNCTION
  WITH ADF/COPFILIN FAMILY PROTEINS (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 11 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE APL FAMILY OF WD-REPEAT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF020054; AAD05042.1; -
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 10.
DR PRINTS: PR00320; GPROTEINBRPT.
DR SMART: SM00320; WD40; 11.
DR PROSITE: PS00678; WD_REPEATS_1; 1.
DR PROSITE: PS50082; WD_REPEATS_2; 5.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
DR Actin-binding; Cytoskeleton; WD repeat.
KW REPEAT
FT REPEAT 58 97 WD 1.
FT REPEAT 102 145 WD 2.
FT REPEAT 146 186 WD 3.
FT REPEAT 189 228 WD 4.
FT REPEAT 234 273 WD 5.
FT REPEAT 320 361 WD 6.
FT REPEAT 365 403 WD 7.
FT REPEAT 445 484 WD 8.
FT REPEAT 489 528 WD 9.
FT REPEAT 532 571 WD 10.
FT REPEAT 576 608 WD 11.
SQ SEQUENCE 609 AA; 66566 MW; 1DB89BE431DEE32A CRC64;

Query Match 2.0%; Score 143.5; DB 1; Length 609;
Best Local Similarity 19.9%; Pred. No. 0.17;
Matches 125; Conservative 92; Mismatches 238; Indels 173; Gaps 32;

OY 65 SGNYLAVTGADYIVKIFD---RHGQRSEINLPNCVAMDMDDKGDVLAIVIA--EKSS 117
DB 71 SGYIASGVDSGKRLIMDTTQKREHLKYEQFPAGKIKOLAMVEDSKRIIAVBEGRKFG 130
OY 118 CIYLDAMNT-----NK-----TSQLDN-----GMRDMSFLL-- 144
DB 131 AVFLMDSGSSVGEITGNKINSVDIKOTPRYLATGSDNDCAAFEGPPEKFKFTLSDH 190
OY 145 -----ASKVGSFLAVGVKGNLKIYNKQTSRKIPVIG--KHKKRITCGGMNENLX 193
DB 191 TRFVNCVRSPPDGNRRPRAASADGQIFLYDGKTEKVCALGGKAHOGGIIAISWSPSSQ 250
OY 194 ALGGE-DKMITVSNQEGDTIRQT-QVRSSEPKNMQFLMKMDRTSAESMISVVLGKKTLL 251
DB 251 LLSASDCKTAKIMDVGANSVSTFNNMGSNVLDDQLGCLMQKDH-----LLSLSL-SGYI 303
OY 252 FFLNLNEPNDPADLEPQDFGNIVCYINW---GDGRIMIGFSCGHEVYVSTHTGELGQET 308
DB 304 NYFDKNNPPKPLRY-IKGHSKSTOCLTVHKNCKSKSYIYSGSNDGHNINWDSPDGE--NDG 360
OY 309 FOARNKHNDNTSTIAVSQGTMLKKVATCGDNCKIKDIDLKAMVYILNDEENKKGITLSWT 368
DB 361 FSGKGHTNOVSRAVDE-MQDLVTC-----SMDD-----TVRYT 393

```

```

OY 369 DDCQLLALSTORGSLLHFLKRLPILG-DACSTRIAYLTSLEVTANPVGEGLPIYVSVD 427
DB 394 -----LSKRDYSGDA-----VKMD 409
OY 428 VEFNEVAVG---LYHLANGMNRAMFYVLGENAVK--KIKMEYLGTVASICLH---SD 478
DB 410 VQFKCLAVGGGVGVVVICIQ-----IYLMKKKKCFADIDGLY--EPEAVAVHPGGGS 461
OY 479 YAAALFEGVQQLHILISELIDAOEERTRLFPVADKCRILCHALNSDLITYGTDGVQ 538
DB 462 VAVGGTGDGNVRLSIQSTLSKSDKTLKAKGPTD-----LAYSIDGAFIACVDANKVVT 516
OY 539 YFYEDMQFVND--YRHPVSVKKIFPDNCTRLVFIDEKSDGEVYC-PVND--ATYEIPD 593
DB 517 VFSVPGGVYEHNVFYGHAAVVCIAISPDEH--FASGMDMVMYVWTVSDEPETRIKIPD 574
OY 594 FSPF--IKGLVEMNPMKGVFIAYDDDD 619
DB 575 AHRLLHVSGLAW---LDEHTLVTTSHD 598

RESULT 15
CLH_CAEEL STANDARD; PRT; 1681 AA.
AC P34574;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Probable clathrin heavy chain.
GN T2065.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Berks M., Smith A.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
  COATED PITS AND VESICLES (BY SIMILARITY).
CC -1- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
  LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT
  (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
  VESICLES (BY SIMILARITY).
CC THE TRISKELION, THIS REGION CONTAINS THE TRIMERIZATION DOMAIN AND
  THE LIGHT-CHAIN BINDING DOMAIN INVOLVED IN THE ASSEMBLY OF THE
  CLATHRIN LATTICE.
CC -1- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z30423; CAA83003.1; -
DR PIR: S42369; S42369.
DR HSSP: P14442; 1BPO.
DR WormRep: T2065.1; CE00480.
DR InterPro: IPR001473; Clathrin_propel.
DR InterPro: IPR000547; Clathrin_repeat.
DR Pfam: PF01394; Clathrin_propel; 7.
DR Pfam: PF00637; Clathrin_repeat; 7.
DR SMART: SM00299; CLH; 7.
KW Hypothetical protein; Coated pits.
SQ SEQUENCE 1681 AA; 191541 MW; 44D15C61339009D9 CRC64;

```

Query Match 2.0%, Score 143.5, DB 1; Length 1681;
 Best Local Similarity 18.1%, Pred. No. 0.74;
 Matches 213; Conservative 158; Mismatches 372; Indels 431; Gaps 57;

```

QY 65 SGNLYAVGADYIVYKIFGRHGGKREINLPGNCVAMDM-----KDGVLAV- 111
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 688 ASKHEQLGADLLIMFENHNSYEGLEYFLGSIVNFSDPEVHEKYIOAATFGOIKEE 747
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 112 -IAKSSCIIYMDANTNKTSQLDNGMDMSFL-----WSKVSFLAVGVKGNIXIYNH 166
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 748 RICRESQC---YDAREVKNFLKEAKLNDQLPLIYCDRHNWHDV-----LYLIRN 796
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 QTSRKIPVLGKHKTRKRTGCGMAENLXALGGEKMIIVSNQSGDTIKOTQVSEPHXNQF 226
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 797 QLOKYEIVFVKV-----NAARLPIVVG--ALLDVDCSE-DAIKOLLINTRG----- 840
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 227 FLMKMDRTSAESMISVILGKKTFLN-----LNEPDNP 262
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 841 ---KFD-----IDELVEEVEKRNRLKLNHMLESKIQEGATDAATHNMAKIYIDSNNP 892
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 263 ADLEFQDFGNIVCYNYMGDGRIMIGFSC---GHFVVISHTGELGOEIFQARN----- 313
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 893 E--RFLKE-----NRYDSKV--VGKCEKRDPRHAFLSYERGGCDALINVCNENSLF 942
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 314 -----HKDNLTSIAVSQTLNKNVATGDMCIKIQ----- 341
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 943 KNLARYLVKRRDFTLMEQVLNEENVHRRQLIDQVYOTALSETODEDISVTYKAFMAADL 1002
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 342 --DLVDLKDMYIIL--LDEENKGLGTLSTWDDGOLLALSTGRSLHVFTRKPLIGDACS 398
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1003 PNEELLELEKTVLDNSAFSEHRNLONL-----ILTKRADRTVMEYIOKLNDYADPI 1057
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 399 TRIAYTSLLEVTVANPVEGELPTVSVDBENFVAVGLYHLAVGMNNRAMEYVLGENAV 458
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1058 ANIAITSELYEEAFA-----IFKRFVNSAINV---LIEVNNLDRAYEFAEK-- 1103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 459 KKLDMETLGTVASICLHSDYAAALFECKVOLHIESILDAQERETRLPRAVDKCR 518
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1104 -----CNQSDWASLAKAQLQONLKEAVDSFIKADPGAYMEVVKCS- 1147
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 519 LCHALTSDFLIYGTDTGVVQFYIEDMOFVNDYRHPVSVKKIFPDNGTRLVFEDEKS-- 576
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1148 -----QTEHMEDLVRYLO--MARKKSRESYIETELVFALAKTGR 1184
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 577 ---DGFVYCPVNDATYETIPDESPTIKGLVWENPMDGVPFAYDDDKVYTYVPHKDTIQ 632
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1185 LTELLEEFYIAGPNHAQIQIGDRC-----FDNGMF----- 1213
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 633 GAKYILAGSTVYPRAHKFLLLYNGELTQOTSGKVNNTIYLSHGFLSNLKDGPDELAPM 692
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1214 -----DSAKILF-----NNV-----SNFAKLSYTLVR-- 1235
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 693 LAHNLMKRFSDAMEMCRILNDEAMNELLARAC-----LHHMEVEFAI 735
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1236 -----LGEYGAVDAAKANKANSTKTWKQVCFSCVNGEFLAQMCGLHIVYNADELEELI 1289
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 736 RYVRRIGNV-GIVNSLEQIKGIEDYNLAGHLAMFTDYNLAODLYLASCPYAL----- 790
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1290 NFYQDRGHFEELLALLEALGLE-----RAHMGCFTE-----LATLYSKY 1329
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 791 ---EMRDLQWMDSALQALAKL--APDQIPFISKEVATOLEFAGDYVNALAHYKGCITGDN 846
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1330 KPEKMRHELLEFWSRVNIPKYLRAEQAHLWS-----ELVFLYD-----KYE----- 1371
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 847 KEHDEACLAGVAQMSIRGDIRRGVNAQALKHPSRYLKRDCGALLEMMKQFSEA--AOLYEX 905
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1372 -EYNNALTL-----MMQHPTESWRE-----OHEKEVIATAKANV 1403
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 906 GLYYDKASVYIRSKNNAKVGLDLPHVSSPKIHLOYAKAKBADGRYKEAVVAAYENAKOMQ 965
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1404 ELYY-KAMQFVLDYKP--LLNDLLLT--VLSPLRD-----HSRTVLEFPNKLQ-I 1447
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 966 SVIRIYI---DHLNPEKAVNIIVRETOSLDGAKKAVARFFLOLDYGAIOFLVMSKCNNE 1022
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

DB 1448 PLVKPYLROVOVNLNN--KAIN-----EALNQLLIDEEDH--AGLRSSIEAQDNFD 1493
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1023 AFTLAQOHNMKEIYADIIGSEDTTNEGYQIALY--FEGEKRYLQGRFLLCGQYSRALK 1081
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1494 NITLQOLEKHPLV-----ERRISAYLFKGNRNKOS---TELC----- 1530
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1082 HFLKCPSEDNV---AIEMALET--VGOAKDELLT 1111
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1531 -----KKDKLYKDAMEYAAESRNGELABELLS 1557
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: May 21, 2002, 14:52:19
 Job time: 158 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 14:49:40 ; Search time 50.02 Seconds
(without alignments)
4779.668 Million cell updates/sec

Title: US-09-729-653-2
Perfect score: 7257
Sequence: 1 HSLIGRCSRLGIDGNNAVC.....AAQLKTSIDCTQYLRTHEEL 1382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4700	64.8	905	4	Q9HCD4 homo sapien
2	2780	38.3	1366	4	Q9V8W2 drosophila
3	2354	32.4	481	4	Q9H5S0 homo sapien
4	1157	15.9	519	5	Q46018 caenorhabd
5	924	12.7	634	5	Q46020 caenorhabd
6	289	4.0	1205	5	Q9W097 drosophila
7	289	4.0	1749	11	Q9JKU3 rictus norv
8	271.5	3.7	1242	4	Q9HBG6 homo sapien
9	270.5	3.7	1758	5	Q22830 caenorhabd
10	268.5	3.7	1292	4	Q9HBG5 homo sapien
11	263	3.6	1198	4	Q9U880 homo sapien
12	263	3.6	1241	4	Q9HAT9 homo sapien
13	256.5	3.5	1462	4	Q60332 homo sapien
14	251.5	3.5	1653	4	Q9UG01 homo sapien
15	248.5	3.4	1462	4	Q96RY7 homo sapien
16	246	3.4	1131	4	Q9BTR9 homo sapien

17	246	3.4	1160	4	Q9BRY4 homo sapien
18	235.5	3.2	993	4	Q9NV68 homo sapien
19	235	3.2	1443	5	Q9VPRO drosophila
20	223.5	3.1	955	4	Q9H7P3 homo sapien
21	218	3.0	764	4	Q9P2H3 homo sapien
22	216.5	3.0	864	4	Q9H9Z0 homo sapien
23	215.5	3.0	705	4	Q96KK0 homo sapien
24	209.5	2.9	1307	5	P90757 caenorhabd
25	206	2.8	821	4	Q9NMR6 homo sapien
26	190.5	2.6	1119	5	Q18859 caenorhabd
27	190	2.6	766	4	Q9P2L0 homo sapien
28	185	2.5	1090	4	Q9ULP1 homo sapien
29	183.5	2.5	588	4	Q96KK1 homo sapien
30	183.5	2.5	656	4	Q9UG52 homo sapien
31	180	2.5	601	11	Q91YV4 mus musculu
32	178	2.5	1047	5	Q45401 caenorhabd
33	174.5	2.4	968	5	Q9V5J6 drosophila
34	171.5	2.4	1319	10	Q9FNA4 arabidopsis
35	170	2.3	887	5	Q9W040 drosophila
36	164	2.3	472	10	Q9FFV8 offif8 arabidopsis
37	164	2.3	493	4	Q9U196 homo sapien
38	164	2.3	493	4	Q75869 homo sapien
39	164	2.3	504	4	Q9ULH8 homo sapien
40	163	2.2	496	4	Q9UM11 homo sapien
41	161.5	2.2	558	4	Q9UFJ9 homo sapien
42	161.5	2.2	911	10	Q9LW87 arabidopsis
43	161.5	2.2	1223	4	Q9U1F5 homo sapien
44	161	2.2	493	11	Q9RIK5 mus musculu
45	161	2.2	493	13	Q42585 xenopus lae

ALIGNMENTS

RESULT 1
ID Q9HCD4 PRELIMINARY: PRT: 905 AA.
AC Q9HCD4:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE KIAA1638 PROTEIN (FRAGMENT).
GN KIAA1638.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirotsawa M., Ohara O.;
RT Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro."
RL DNA Res. 7:273-281(2000).
DR EMBL: AB046858; BAB13464.1;
DR InterPro: IPR000547; Clathrin_repeat.
DR SMART: SM00299; CLH; 1.
FT NON_TER
SO SEQUENCE 905 AA; 102450 MW; FD370996F00FE1BF CRC64;

Query Match 64.88; Score 4700; DB 4; Length 905;
Best Local Similarity 99.88; Pred. No. 0;
Matches 903; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 375 ALSTQRSGLVFLFKLPILGDACSTRVAYLSLETVANPVEGELPTVSVDEPFVA 434
Db 1 ALSTQRSGLVFLFKLPILGDACSTRVAYLSLETVANPVEGELPTVSVDEPFVA 60
Qy 435 VGLYHLAVGNMNRAMPVYLGDNNAVKRLKDEYLGTVASICLHSDYAAALEGKVOYLIE 494
Db 61 VGLYHLAVGNMNRAMPVYLGDNNAVKRLKDEYLGTVASICLHSDYAAALEGKVOYLIE 120

```

OY 495 SEILDAOEERERTRLPVADDKRILCHALTSPLLYGTDTGVVQYFIEDMOFVANDYRHP 554
    |||||
Db 121 SEILDAOEERERTRLPVADDKRILCHALTSPLLYGTDTGVVQYFIEDMOFVANDYRHP 180
OY 555 VSVKRIPEPNCNTRILVFIKESDGFVYCCVNDATYEIPFSPITICVLMENPMKGVYI 614
    |||||
Db 181 VSVKRIPEPNCNTRILVFIKESDGFVYCCVNDATYEIPFSPITICVLMENPMKGVYI 240
OY 615 AYDDKRVYTYVHFKDTIOGAKVILAGSTKVPFAHKPRLLYNSELNCOFOSGVNNIYST 674
    |||||
Db 241 AYDDKRVYTYVHFKDTIOGAKVILAGSTKVPFAHKPRLLYNSELNCOFOSGVNNIYST 300
OY 675 HGFLESLNKDXGDELPRMPLAHMLMKRESDEAMKCRILNDEAMNMLARACHHMEVEFA 734
    |||||
Db 301 HGFLESLNKDXGDELPRMPLAHMLMKRESDEAMKCRILNDEAMNMLARACHHMEVEFA 360
OY 735 IRVYRIRGVNGIVMSLEOIKGIEDYINLAGHLMFENDYNLADLYLASSCPILALEMR 794
    |||||
Db 361 IRVYRIRGVNGIVMSLEOIKGIEDYINLAGHLMFENDYNLADLYLASSCPILALEMR 420
OY 795 DLQWDSALQALAKHLAPDQIPEISKEYATOLEFAGDYVNALAHYEKGITGDKHEDEACT 854
    |||||
Db 421 DLQWDSALQALAKHLAPDQIPEISKEYATOLEFAGDYVNALAHYEKGITGDKHEDEACT 480
OY 855 AGVAMSTRMGDIRRGVNAKHKHPSRVLLKRCOGATLENNKOFSEAAOLYEKLYYDKAAS 914
    |||||
Db 481 AGVAMSTRMGDIRRGVNAKHKHPSRVLLKRCOGATLENNKOFSEAAOLYEKLYYDKAAS 540
OY 915 VIIRSKNNAKVGDLLPHVSSPKIHLQYAKAKADGRYKAAVAVAYENAKOMOSVIRIYD 974
    |||||
Db 541 VIIRSKNNAKVGDLLPHVSSPKIHLQYAKAKADGRYKAAVAVAYENAKOMOSVIRIYD 600
OY 975 LNPEKAVNIYRETOSLDGAKAVANFLOLDGYSAIOFLVNSKCNNEAFTLAQOHKME 1034
    |||||
Db 601 LNPEKAVNIYRETOSLDGAKAVANFLOLDGYSAIOFLVNSKCNNEAFTLAQOHKME 660
OY 1035 IYADITIGSEDTNEDYOSATLFEBEKRYLQAGKFFLLCGOVSRLKHFLLKCPSSSEDNA 1094
    |||||
Db 661 IYADITIGSEDTNEDYOSATLFEBEKRYLQAGKFFLLCGOVSRLKHFLLKCPSSSEDNA 720
OY 1095 IEMATEYGOAKDELITJNOLDHLLGENDGMPKDAKYLFRILYMAKQYEAQOTIITAR 1154
    |||||
Db 721 IEMATEYGOAKDELITJNOLDHLLGENDGMPKDAKYLFRILYMAKQYEAQOTIITAR 780
OY 1155 EEOGAGNTRNADVLFSMTAEKSKIKIPSEMATNMLILHSYILVKIHVKNGDHMKGAR 1214
    |||||
Db 781 EEOGAGNTRNADVLFSMTAEKSKIKIPSEMATNMLILHSYILVKIHVKNGDHMKGAR 840
OY 1215 MLIRVANNISKEPSSHIVPLITSTVECHRAGLKNASFAAMLMRPEYRSKIDAKYKKKI 1274
    |||||
Db 841 MLIRVANNISKEPSSHIVPLITSTVECHRAGLKNASFAAMLMRPEYRSKIDAKYKKKI 900
OY 1275 EGMVR 1279
    ||||
Db 901 EGMVR 905

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Branton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Burton R.C., Rogers Y.-H.C., Blazer R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Abmayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Broststein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jitali M., Katush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kolp D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Paule J.M.,
RA Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheibel F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirskas R., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weissstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Glids R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
CC Science 287:2185-2195(2000).
CC -I- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AE003795; AAF57545.1; -
DR FlyBase: FBgn0034452; CG11237.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 6.
DR SMART: SM00320; WD40; 4.
DR Repeat: MD repeat.
SQ SEQUENCE 1366 AA; 153505 MW; E9EBDC9C61AFBBBD CRC64;

```

Query Match 38.3%, Score 2780; DB 5; Length 1366;
 Best Local Similarity 40.8%; Pred. No. 1.2e-182;
 Matches 553; Conservative 286; Mismatches 468; Indels 50; Gaps 14;

```

OY 57 IQFAMOKTSGNLAATGADYIYKIFDRHGQRSEINLPGNCVYAMDMDGDVLAIAKRS 116
    ||||
Db 19 YEFIWOKA---LATTGTGDSVALYNROGOLVORITLISGLSGFAMDEGDGLITTSGS 75
OY 117 SCIYLMADANTFKTSOLDGMDRMDOMSFLLMSKVGSPFAGTVAGNLIXYHGT-SKIPVL 175
    ||||
Db 76 PNITLMDYNSOEKISYERGLNDPLTCLIMSKOQOLAVGTGNGALNIYHSSGKRPYVL 135
OY 176 GKHTKRITCGCWNNAENLXALGEGEDKMTVSNQEGDTIROTVRESEPNXMKQFFLMKMDRT 235
    ||||
Db 136 GKHSKRITCGAMSAQNLNLALGSEDEKSFSLSNEDGDTVAVVOLRADPDTMYFAEMANDERI 195
OY 236 SAAESMISVNLGKTLFLPLNLNPNPNADLEQDFGNIVCYNMGDGRIMGSCGFY 295
    ||||
Db 196 -AGDAIMSIITGKRLFLYLPPEPNPELGFQSRYSGLIMOHKMGWDYILLLGSHYV 254
OY 296 VISTHTGELGOEIFOARHKNLNTSIAVSQTLNKAATGCDNCKIKODLVLDKDMVILNL 355
    ||||

```

Db 255 A1STHPRVGOELMOVKNHKSJLTGLAYCPTLIDIVASCDDSIKHSITNLOJETERITV 314
 QY 356 DEBNKGJLSTWTDGOLLALSTORGSLHVELKLPILDGACSTRALATLSLELVY--- 412
 Db 315 -PBAAGVQMDWSPDQOLLAVTNGVYIYVTKLPPLRLVSPARIVLSSLAESVSYV 373
 QY 413 ANVEGELPTVSVDEPNVAVGLYHLAVGNMRAFFVYLGEM---AVKELDMETIGT 469
 Db 374 APKTSLSLPRRLPLEBEPPTMAVGNFATGIEKHWFIDGKSLGEERPLSERPPRS 433
 QY 470 VASICLHSDYAALFEKGVOLHLESEILDQOERET---RLFPVADCKRILCHALTSD 526
 Db 434 VESMKINADYACALCPRLQILQIAADNPCKDKLOAVFETPALNPSDVAIVICFALSQ 493
 QY 527 FLTYGDTGVVOYFYIEDMGFVNDYRHPVSKVIFPDNSTRVLVFIKESDGVYVCND 586
 Db 494 LTFADIDGLHVLVYSLKMDSCITYRHSWGIROLFMDIBETKVFIDHDSQGVFLPYVE 553
 QY 587 ATVEIPDFPTLIGVLMENMPDKGVFIADVDDKVYTYVFNKDTIOGAKVILGYSFVKPF 646
 Db 554 EALLIDIPKOCGLMD--LTQPNIFISDARIIVTHVVRHSVQGTHTLVMGESKLP 611
 QY 647 AHKPLLYNGELTQOTQSGKVNVIYSTHGFSLNKDXGPEDELRPLAHNLMLKRSDDA 706
 Db 612 GQFPLLCGEMALHIDGQYATQSLSTH--VYVPSNSQANLQML---LKLRYNDEAY 665
 QY 707 EMGRILNDEAANMLRACGHNHEFAITRVYRIGVNGVMSLEQIKGIEDYNLLAGHL 766
 Db 666 KICKONHOSAMREFEQALSDLEPDLATRAYQSLGMAALVNLSELRYVEDLMLGCC 725
 QY 767 AMFTNDYNLADLYLASSCPRIALLENMRDLOHMSALQIALKHLAPDOIPREISKEAYOLE 826
 Db 726 CTELAGYDQAKEMHMGVYRRAALDLCDRLQDQALLHANKNDQEVPIAREVYQOLE 785
 QY 827 FAGDYNALAHYKGTIGD-----MKHEDACLAGVAMQSTIRMGDIR 868
 Db 786 FGNVYTDALHYEGYKEDINSKETETDALMDSPEYEHVIRCKGIGIARTISIRAGDFR 845
 QY 869 RGNVALKHSRYLKRDCGAILENMKQFSEAOLYEGKGLYDKAASYIASKMMAKVGDL 928
 Db 846 RGTQYVLEEDQOLDFDCABELTAVGHLEAGLYERGFDECGHYIALKMKNNKNNI 905
 QY 929 LPHVSSPKIHLQYAKAKADAGRYKAVAVAYENAKOMOSVIRIYLDHLENPEKAVNIRET 988
 Db 906 LPRVSKTKLHAAYAKAKENGHYEAIIRSYRIGDGDACRITLYLDHLCODHAASEIYLES 965
 QY 989 OSUDGAKMVARFLOLDIGSALQIPLVMSKCNNEAFTLAOQHNMKEIYADIGS-EDTNN 1047
 Db 966 RSMDSAKLLAKFYOKIGDVEQALQFLVICGVEEAFALORHNKLRHNGELLERYENAKS 1025
 QY 1048 EDVOSIALYEEGKRYLQAGKFFLLCGOVSRAKHLKPCP--SSEDNVVAIEMAIETVGOA 1105
 Db 1026 SDTLALAHFEGEKRYTLACKYFLAREFYKALRFLKASAFNNEBOVSLSLADICVATS 1085
 QY 1106 KDELLTNQILDHLENGDGNPKDAKYLFLYMLAKOYREAAQTAIITAREOSAGNYRNA 1165
 Db 1086 NMQGLATQLEFLLEGDGVKDPKRYFLRYMLARKHYKAKDAKTAVALIANEQOAGNKA 1145
 QY 1166 HDVLFKSYAELAKOKIKISEMATNMLIHSYLVKTIHVNGDHMKCARMLIRYANNISK 1225
 Db 1146 RDLVSYVOELRNRNLSVTRKEMRHOFILRHVLRHVNHLHLAKLALVOYACISQ 1205
 QY 1226 FPHSHVPLSTVYIECHRAKLSAFSAFAMLMRPREYRSIDAKYKKIKGIVRRP-- 1282
 Db 1206 FPHSHVPLSTVYIECHRAKLSAFSAFAMLMRPREYRSIDAKYKKIKGIVRRP-- 1282
 QY 1283 ---ISEIEAATPCPFCKFLLPCELCPCCKNSIPYCIATGRHMLMDWTVCPCDFPA 1339
 Db 1266 KOLRDEIDGETMBCPIODSNLANNEVTCYSCKTTPLICATGQIILIKQWLTSCQOQDFLC 1325
 QY 1340 LYSELKATMLNTESTCPKCSERLNAOLKISDCTQYL 1376
 Db 1326 FRAEMENILSENGECPMGCGENVAPEOLLDAVEDIRPYI 1362

RESULT 3
 ID 09H550 PRELIMINARY; PRT; 481 AA.
 AC 09H550;
 DT 01-MAR-2001 (Tremblrel, 16, Created)
 DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
 DE CDNA: FLJ23127 FIS, CLONE LMG08350.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Oro T.,
 RA Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y., Ota T., Suzuki Y.,
 RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
 RA Isogai T., Sugano S.;
 RT NEDO human cDNA sequencing project.*;
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 3 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AK026780; BAB15550.1;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 3.
 DR SMART: SM00320; WD40; 3.
 KW Repeat; WD repeat.
 SQ SEQUENCE 481 AA; 53542 MW; B1C7AE096581F02 CRC64;

Query Match 32.4%; Score 2354; DB 4; Length 481;
 Best Local Similarity 99.3%; Pred. No. 5,5e-154;
 Matches 449; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 41 MKRIFSLLEKTYWLGAPLOFAMQKTSNGYLAVTGADYVVKIFDRHGRSEINLPGNCVAM 100
 Db 1 MKRIFSLLEKTYWLGAPLOFAMQKTSNGYLAVTGADYVVKIFDRHGRSEINLPGNCVAM 60
 QY 101 DMDKGDVLAVIAEKSSCIYLDANTNKTSQLDNGMRDMSFLMSKVSFLAVGTVKN 160
 Db 61 DMDKGDVLAVIAEKSSCIYLDANTNKTSQLDNGMRDMSFLMSKVSFLAVGTVKN 120
 QY 161 LXTYNNQTSKRTYVLYGKHTRTITCGCNAENLXALGDEDMATVSNQEGTITQTVRSE 220
 Db 121 LXTYNNQTSKRTYVLYGKHTRTITCGCNAENLXALGDEDMATVSNQEGTITQTVRSE 180
 QY 221 PXNMQFFLMKMDRTSAEEMISVVLGKKTLLFNLNEPNPADLEFOQDFGNIVCYNWY 280
 Db 181 PXNMQFFLMKMDRTSAEEMISVVLGKKTLLFNLNEPNPADLEFOQDFGNIVCYNWY 240
 QY 281 GDRIMIGFSGHGFVISTHTGELGOEIFQARRHKNLTSIAVSQTLNKVATCGDNCKI 340
 Db 241 GDRIMIGFSGHGFVISTHTGELGOEIFQARRHKNLTSIAVSQTLNKVATCGDNCKI 300
 QY 341 ODVLVDKDMVYIINLDENGGLGTLSTWDGQLALSTORGSLHVELTKLPILGDCSTR 400
 Db 301 ODVLVDKDMVYIINLDENGGLGTLSTWDGQLALSTORGSLHVELTKLPILGDCSTR 360
 QY 401 IAYTSLLEVTVANPVGELPTIVSVDEPNFAVGLYHLAVGNMRAFFVYLGEMAVKK 460
 Db 361 IAYTSLLEVTVANPVGELPTIVSVDEPNFAVGLYHLAVGNMRAFFVYLGEMAVKK 420
 QY 461 LKDMYLGTVASICLHSDYAALFEKGVOLH 492
 Db 421 LKDMYLGTVASICLHSDYAALFEKGVOLH 452
 RESULT 4
 ID 046018 PRELIMINARY; PRT; 519 AA.

AC 046018; 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DE 2K520.1 PROTEIN.
 GN 2K520.1
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Steward C.A.;
 RL Submitted (Mar-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RT none;
 RA "Genome sequence of the nematode C.elegans: A platform for
 investigating biology."
 RT Science 282:2012-2018(1998).
 DR EMBL; 292822; CAB07299.1;
 DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
 SQ SEQUENCE 519 AA; 59152 MW; 38F80CA32D06E02D CRC64;

Query Match 15.9%; Score 1157; DB 5; Length 519;
 Best Local Similarity 45.0%; Pred. No. 2.3e-71;
 Matches 230; Conservative 114; Mismatches 157; Indels 10; Gaps 4;

OY 724 AGLHMEVFATRYVRICNVGLVSLDQIKGIEDYNLLAGHMTNTNYNAODIYLAS 783
 DB 2 AALLSDVMAKIFREICDAAVTALEITEENKLLHAQYITLLSRDAEOLYLES 61
 OY 784 SCPIALEMRRLDQHSALQALHLPDQIPFISKEVAIQLFADYVNAALAHYEKG- 842
 DB 62 SRPMELNNRRDLEPRKALVLAETMPKREIPYLSKEVQOELELGDHNSLANYEKGVM 121
 OY 843 -----TGDKHEDEACLAAGVQNSTMGDIRGVNOLKHPKRVYLRDCCAILLENKQPS 897
 DB 122 ENPOMLPFLOEHNEICOSGIARMAIKTGDLRRGVOLAKOLEGRVVRDCAIILEQMKQYT 181
 OY 898 EAAOLYERGLYDKAASVYIRSKNNAKVGDLPHVSSPTIHLQYAKKADGKRAVVA 957
 DB 182 EAAQLYEVGLFDRRAAVLCKANNAKVGELLDHVKSPKTHIQYKIMKEKKYKAAVVC 241
 OY 958 YENAKOMQSVIRIYDLHNNPEKAVNIVETQSLDGAKVAFELQDGYSAIOFLVMS 1017
 DB 242 YETGRDYDQVRLDPLNDPEAVVRESRSIEGAKLVAKFVYKLDGYNASIOFLVMS 301
 OY 1018 KKNNEAFTLAQOHKKEITADIIGSEDITNEDYQSIALYFESEKRLQAGKFFLLCGYS 1077
 DB 302 QCVQAEFFELAEKNNAVREYAKAIEQHNISQALE-LAEYYNVNDMFMAKFFYTOAGOVN 360
 OY 1078 RALKHFLCPSEEDNVAIEMATEVQAGDELITNOLIDLHLENDGPKMDKYLFLRM 1137
 DB 361 NAINLFLK--KDDDENCVALAVDCGKSKDKTLNKLKLVFLGFE-DGNVYKDAQFLRLV 417
 OY 1138 ALKOYREAAQTAAIILAREBOSAGNYRNAHDVFSMYAEKLSOKIKIPSEMATNLMLHSY 1197
 DB 418 GLGRFKDAQGTAVVAQIHQAGKGNRIARDLFLQMHQGLREKMMRPLDMNNSIMAHISHY 477
 OY 1198 ILVKIHKNGDHMKGRMLIRVANNISKFPS 1228
 DB 478 IIVKALINRKETLLAARLLIFTCGEIQRPT 508

RESULT 5
 O46020 PRELIMINARY; PRT; 634 AA.
 AC 046020;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE 2K520.3 PROTEIN.
 GN 2K520.3
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Steward C.A.;
 RL Submitted (Mar-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RT none;
 RA "Genome sequence of the nematode C.elegans: A platform for
 investigating biology."
 RT Science 282:2012-2018(1998).
 CC -1- SIMILARITY: CONTAINS 4 WD_REPEATS (TRP-ASP DOMAINS).
 DR EMBL; 292822; CAB07301.1;
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 4.
 DR SMART; SM00320; WD40; 3.
 DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 634 AA; 69905 MW; F35637A84B16930A CRC64;

Query Match 12.7%; Score 924; DB 5; Length 634;
 Best Local Similarity 30.4%; Pred. No. 3.7e-55;
 Matches 204; Conservative 130; Mismatches 239; Indels 98; Gaps 8;

OY 56 PIQFAMQKTSNYLVAVTGADYIVKIFDRHQRSEINLPGNCVAMDWDKGVAVIAEK 115
 DB 37 PIHHR-RENGTAVACANNTVIYDKKGNVIDALNPTGKLIDIAWKEGDVLAIVAN 95
 OY 116 SSCITLMDANTKTKTOLDNG---MRDQMSFLMSKVGSLAAGYKNGNIXIYNHOTSRI 172
 DB 96 TGTITLMDVNSKNTPTVESGATSSKELPTCLAMSPSTPLVIGNNGNIVNHTSRRI 155
 OY 173 PVLGKHTKRTGCGNAENMLXALGDEKMITVSNQGDITIROTVRSEPNQFELMKMD 232
 DB 156 AMGKHQRSVQITVTPBEDYVISCDDNTLSVTLEGTVSTTTNGEPTNNDY--GSVN 213
 OY 233 DRTSAESMISVYLKTKTLFELNLNPEPNADLEFQDQGNIVCNWGDGIMYGFSCG 292
 DB 214 GGGGQVTVMSVYIKITIMLAHYNALDEPVNLQOEKGNHSHRWFDGILLGFPRG 273
 OY 293 HGVVISTHGEIGELFQARNHKNLTSIAVSOTLNKVAATCGDNCIKIODLVLDKMYI 352
 DB 274 YIISISAHNNEIGSELVSLETKGLASIAVSTSENKLLITIGDNVKKVDELITVIML 333
 OY 353 LNLDEKNGKGLTSMWDDQLALSTORGSLHVFLLKPLILGDASTRIAYLTSLEEVTV 412
 DB 334 TELERE-KNLSEIETVEDQGLAIVASQSGVLISFVTKMPTLAASVNSICVLTNLQTV 392
 OY 413 ANPVGELPIYVSUVEPNEFAVAGLYHLAVGNMNNRWY-----VLGNNAV 458
 DB 393 VAEEVKKGSSTLELIEPTVMGLGPLNLAVANNNTVFFEDYHTPAAQMAAOLQSTQSA 452
 OY 459 KK-----LKDMYLGLTVASICLHSDYAAALFEGKYOLHIESEIIDAQERETRLFP 510
 DB 453 EKPITVAEPINRVEYLSVTNLIQNLNMAAANFSGRLRLHRIIN-----SDNVSIIEP 507
 OY 511 AVDDCRILCHALTSDFLITVGTGVOGYFYTEDQFVNDYRHPVSVKIFPDNGTRLV 570
 DB 508 EARNRATLXVALTENFL----- 526
 OY 571 FIDKSDGFCVPCVNDATAYELPDSPITKGLVEMWPMKGVFIAYDDKVTYVYFHKDT 630
 DB 527 -----FTTSYKGAIMEFTTIDKNFAVFDSONIVFLLSKOH 563
 OY 631 IGGAVIILAGSTKVPFAKHPILLVNGELTLCQOSQKVNINVIYSTGFLSNLKDQXPDEL 690


```

Db 764 YGEFEAEKLYIDADRDLAIELRMTLCDWFRVVDLYRMGGSGVSDQOMELMREIGHNF 823
QY 810 APDQIPFISKEAIOLEFAGDYVNLALAHYEKGTGDNKEHDEACLAGVAMQSIKRMQDIR 869
Db 824 ANLRSMESAREYKESHLYEGYMEALYHLE-----QDDLEK 860
QY 870 GYNQALNHPKRYLKRDCGAILNKKOFSEDAOLYKGLYDKAASV--YISKNNAKYGD 927
Db 861 CV-EKRLPEKSPLLPK-LAEMLASVCMSEAVQAHLR--FGDQKAAVATCVNLRQMGFAVE 916
QY 928 LLPHYSSPKIHQYAKAKE--ADGRYKAAVVAAYENAKOMQSVIRIYLDHLNPKKAVI 984
Db 917 LAQRFQLPQVOTLLAKHAAQLQEBRLEKEAIEQMRA----- 953
QY 985 VRETQSLDCAKAVAFRFLQIDGYSALQFLVMSKCNNEAFTLAQOONKMEIYADIISPD 1044
Db 954 ---GRLHDAARLSQ--MAERQEKRAPLLRIKKLYLAALLAEHLK-----AVA 999
QY 1045 TTNEDYOSIALYEESEKRYLQAGKFFLLCGOYSRLAKHFLKCPSSSEDVNAIMAEIVGQ 1104
Db 1000 TTEIDYAS-----GRNTLL-----DSIALE----- 1019
QY 1105 AKDELLTQQLDHLLEGENDMKDAKYFLRYMALKOYREAAQTALITAREQASAGNYRN 1164
Db 1020 -----DAALERLWHCAEAYH-----PMLLAQRLQRLFG----- 1047
QY 1165 AHVDLFMYAEKLSOKIRIPSEMATNLMLISYLVKIHVKNQDMKGAIRLVANNIS 1224
Db 1048 -----IVHSAAVTVAVRLRYEDVLP----- 1067
QY 1225 KPPSHIVILSTVIECHNAGLKNASFSAAMLKRPETRSKID---AKYKKKIEGMVRR 1280
Db 1068 --PEHIVSYLLASCADRAFGTCSKAF-----MKLEODANLRPEATLQRYEELAGIYFAK 1119
QY 1281 PDISEIEATPPCRPFCKFLPPECCELLCPGCKNSIPYCIATGNNHMK---DDW--TVCPHC 1335
Db 1120 YDEPBTGDRVDCYSCGYVPDSSPCBECNARFACISSGKPTIOPTNINICTTCHHC 1179
QY 1336 DEPALYSEIKIMLNTSESTPMC 1357
Db 1180 AAPT-----EISRHTCSPIC 1194

```

```

RESULT 7
Q9JUK3 PRELIMINARY; PRT; 1749 AA.
ID Q9JUK3
AC Q9JUK3
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE SELECTIVE LIM BINDING FACTOR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20250882; PubMed=10788441;
RA Howard P.W., Maurer R.A.;
RT "Identification of a conserved protein that interacts with specific
RT LIM homeodomain transcription factors.";
RL J. Biol. Chem. 275:13336-13342(2000).
CC EMBL: AF226993; AAF68274.1;
CC EMBL: AF226993; AAF68274.1;
DR InterPro: IPR002106; AA_trna_ligase_II.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 5.
DR SMART: SM00320; WD40; 4.
DR PROSITE: PS00339; AA_trna_ligase_II_2; UNKNOWN_1.
KW Repeat; WD repeat.
SQ SEQUENCE 1749 AA; 197601 MW; B72C79207F5C0A5 CRC64;

```

Query Match 4.0%; Score 289; DB 11; Length 1749;
 Best Local Similarity 19.1%; Pred. No. 1.2e-10;
 Matches 279; Conservative 195; Mismatches 480; Indels 504; Gaps 70;

```

QY 39 VEMKRIFFILEETWIGAPQ-FANQKTSQNYLAIVGADYIVAFPRHCKQSEINL-PGN 96
Db 1 MQLKHLRTLLSPQDAKAKYTCMAWSQNNAKF-AVCTVDRVILLYDEHGRRRKFTSKP 59
QY 97 -----CVAMDMDKDDVDLAVIAEKSSCIYLDAN-----TKTKSQLNGMD 138
Db 60 MKYGRKSYVKKMAKSPDSTKTA-IGQDNTIYVKKIGDMQDKVYCNKFTQ-----TS 113
QY 139 QNSFLMSKVGSFLAVGVYKGLXYNHQTSKRIPVLG--KHTRKPIYGC----- 186
Db 114 AVTCLQW-PAEYIIVEGLAEGKVRILANRKTNSSTIYGDYSVALVTNCSGKILSGHA 172
QY 187 -----WANEKALGDEKMTIVSQDEDTIRQ 214
Db 173 DGTIVRYEFDDESGSESGKLVNHCPPYALAW-ATNSIVAAGCRRIVAYGKEGHVLTQ 231
QY 215 TQVRESEPKNMFLLKMDRSTSAESM--ISVVLGK-KTLFPLNLTN-----EPDNPADL 265
Db 223 FQYSRDPQEREF-----TTAASPGQSVLSYDRLRFPNNSPRSTIEKAPKEI 283
QY 266 EFQDQGNIVCYNMYGDG-RIMIGFSCGHFVVISHTGELGOEIFQA-----RNHKDNL 318
Db 284 ---ANLVTYVALAMKRDSSRLCAGTLCG-----GVEQPDCLRSIRYKKEEL 328
QY 319 TSIAVSQTLNKATAGDCDKIKQDVLVDKMY---VILNDEENKGLG---TSLWTDG 371
Db 329 TYVGSQILVKNLSSGTRVY-----LKSNGYEV-----EVALIKGERLYVAHTST 376
QY 372 QLLA-LSTQSGSLHFLTKPLPLGDACSTRIVALTSLLEVTVANVEGELPTVSVDER 430
Db 377 LLLGLDLNLR-----LSLEIPWQSGGNGEKYF-----ENE----- 406
QY 431 NFVAVGLYHIAVGNNRRAMFYVLGENAYKKLMDMEYLGTVASICLHSDYAAALFEGKYOL 490
Db 407 -----NYCMIFNAGELTVEGNSDLSVTRTEFNPN----- 438
QY 491 HLIESEILDQAE---ERETRLPRAVDKCRILCHALTSDFLYGDTGVGVGYIEDNQF 547
Db 439 HLISVIRNECRQGMEDKKLAVLDIKTIALVD-----LIGGYNIGTISHESRYDME 492
QY 548 VNDYRHPYVKKIFPDPRNGTRLVLFIDEKSDGFFVCPVNDATYEIPDFSPT--IKGVLMEN 605
Db 493 LNE-----TGHKLLFRDRK-----LRLHLVDISCSKTMILNFCSTVQ 530
QY 606 WPMKGVFLAYD-----DDKYTYVYFHKDTI---QGAK---VILAGSTKYPF 646
Db 531 WPGSDVLAQNRNSLCWYNIAPERYTMSIRGDVGLERGSGKTEVMYEGVTTVAY 590
QY 647 AHKPLLLNGELTQCTQSGKVNIIYLSHGFSLNKKDGPDLRLMLAHNLKRFSDAM 706
Db 591 T-----LDEGLIEFTALDDGN--YTRATAFLETL-----EMTP----- 622
QY 707 EWCRIILNDEAAMNELLARACLHHEVEFAIRYVRIGNVGVLSBOINGI----- 756
Db 623 -----ETFAAMKTLNLSIALEARQLHTAERCSALGHVAKAFLEHETNDAQVREYGG 676
QY 737 --EDYNLAGHLAMFTNDYNAQDLYLASSCPITALLEMRRLQHHNDSALQIAKHLAPDI 814
Db 677 ECTDYYQVRARILAMEKYNKLAEMIFLCONAVEEAMDYOEILHRWDECIJAVEAGHPAL 736
QY 815 PLSKEVAIOL-----EFAGDYVVALAHYEKGLTGDKKHEDEACLAGVAMS 861
Db 737 EKLARDYIQRLLMDIQOERAGELDESQDGLAIAIYLK-----AGLPKA 782
QY 862 IRMGDIRGVNQALNHPKRYLKRDCGAILNKKOFSEAAOLYKGLYDKAASVYIRSKN 921
Db 783 ARL-----VLTRE--ELLANTELVEHITTLALIKGELTERA----- 815

```



```

QY 922 WAKVDDLPHVSPKRIHQ-----YAKAKE-ADGRYKEAVAYENAKOMOSVIRIYLD 973
DB 816 ----GDLFEKIRINFORALBECYCKGNAMKAVELARLAFPEVAVRLEEA--WG-----D 862
QY 974 HLNPEKAVINIVRETQSLODCAKKNVAFFLQDGYGSAIOFLVMSKCNNEAF--TLAQOHN 1031
DB 863 HL-----VQOKO-LD-----MAINHYIEAKRSIKAIEMALGAROW 896
QY 1032 KMEIYADIIGSEDTNEDYOSIA-----LYEEGK-----RYLOAGKFL 1071
DB 897 KKAIIYIDDDKRTASKYTYRVAAOHVASLOEYELAEYTKGDDTKADIMTYOAGR--- 953
QY 1072 LCGOYRSLAHPLKCPSSSEDNVAIEMAIETVGOAKDELTLNOLIDHLLGENDGMPKDAKY 1131
DB 954 ----WEGAHLKAKMCPED-----VSULTYTOAQE-----MEKGGKY 987
QY 1132 --LEFRL-----MALKQYRE-----AAQTALIAEEOGAGYRN 1164
DB 988 REAERLVYVEEEDPLATMKRKHLXYDMIRLVGKNHPRDLSPTHLGLKELETEGRLOE 1047
QY 1165 AHDLFESMVAELKSOKIKIPSEM-ATNLMILHSYILVKIHYKNGDHMK-----GAR 1214
DB 1048 AE-----YHLEAOEKKAVANNMTRSSGLWEBAIRVAKAHGANAHHVAYLMAKSLGGE 1101
QY 1215 MLIRVANNISKFPSPHIVPILSTFVIECHRAGLKNSAFSFAAMLMPREYRSK---IDAKYK 1271
DB 1102 AAVRLNKLKG-----LLEAAID-HAA--DNCSFEFELSRLAKKHTPEIHILRYA 1149
QY 1272 KTEGWAYRRPDISEIEEA 1289
DB 1150 MYLE-----DEGKFEEA 1161

RESULT 8
QY 09HBG6 PRELIMINARY; PRT; 1242 AA.
AC 09HBG6:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE WDRIOP.
GN WDRIOP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21134653; PubMed-11242542;
RA Gross C., De Baere E., Lo A., Chang W., Messiaen L.;
RT "Cloning and characterization of human WDRIOP, a novel gene located at
RT 3q21 encoding a WD-repeat protein that is highly expressed in
RT pituitary and testis";
RL DNA Cell Biol. 20:41-52(2001).
CC - SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
CC EMBL: AF244930; AAG15427.1;
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40: 6.
DR SMART: SM00320; WD40: 7.
DR PROSITE: PS50082; WD_REPEATS_2, 1.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KM Repeat; WD repeat.
SQ SEQUENCE 1242 AA; 141968 MW; 40C1DE7BF34D5E8 CRC64;

```

Query Match 3.78; Score 271.5; DB 4; Length 1242;
 Best Local Similarity 18.4%; Pred. No. 1.le-09;
 Matches 247; Conservative 196; Mismatches 505; Indels 397; Gaps 54;

```

QY 116 SSCITLMDANTNKTSDLDNGRDMQSFLLMSKVSFLAVGVGNLXIYHQTSRKLPV- 174
DB 113 SSDFGLMSPOKSVSK--HKSSSKIIICCSWTDGQYALQAFNGIISIRKNGEKKVIE 170

```

```

QY 175 -LGHKTRITCGCM-----NAENLXALGEDKMI-----TVSNOEG 209
DB 171 RRGSGSLPISISICMNNSSRMESFMNRENDA---EDVIYKRIQIPELPTLSKAVYSSOG 227
QY 210 DTRQGVQVRESEPNMOFFLKKMDRTSAESMISYV-LGKTLFF-LINLEPNDALEF 267
DB 228 SEAEERPEEEDDSPR-----DNLEERNDLIADVDMQKVAFYDLSGQKQTKDRALNF 281
QY 268 QDQPCNVICWYNGDGR-IMIGSCGHFVYISHTGELGELTQARNHNDLSTIVASQT 326
DB 282 DP-----CCISYPTKEVYILG--GSDKQVSLFT-----KGV----- 312
QY 327 LNKVATCGDNCIKIODELVLDKMYVILNDEENKGLTISWTDGOLALSTORGSLHVF 386
DB 313 --RLGTVG-----EQMSWITQCAKSDSNVYVGGCGDTISFY 348
QY 387 LTKLPILGDACSPRIAYLSLEVTYANPYEGELPTVSVDENPNVAVGLYHLAGMNN 446
DB 349 QLIESTVHGILKDRYAVRDSMTDVIVQHLLIT-EQKVIRICKELVKKIAIYRNRLAIOLPE 407
QY 447 RAMEFVYLGEMAAVKKIDMEYLGIVASICLSHDYAAALFECKVOLHLIESIIDAQER-E 505
DB 408 KILIELYS--EDLSDMY-----RVKEKTIKKFECNL-LVVCANHILCOEKRLQ 455
QY 506 TRLEPAVDK-----CRILCHALTSDFLYGTGDTGVQYFYEDMQFVNDYRHPV 555
DB 456 CLSEFSYKEREMQESLIRIKYIGRPRGEGILVGLKNOILIKITVDNLFAYLLKQAT 515
QY 556 SVKRIEPPDNGTRIVFIDEKSDGFVYCPVNDATYEIIPDSPTIKGLVLENMMDKGVFI- 614
DB 516 AVRLCDMSASRKKLAVVENDOTCLVY--DIDTKELLFOGPNANVSAMNTQCEDMLCFSG 572
QY 615 -AYDDKVVYTVFHKDIOGAKVILAGSTKVPRAHNPILLYNG-ELTCQOSQKVVNNIYL 672
DB 573 GGLNTIKASTFVYHROKLG--FVVG-----YNSKILFC----- 604
QY 673 STHGFLSNLKDQXGDELRPLAHNLMLKRPDAMEMCRILINDAANMELARACLIHMEVE 732
DB 605 -LHVF--SIAVEVPOAPRYQY-LDKKILKRAVQIACLOVPTDIRELAMELEGDIFE 660
QY 733 PAIRVYRRIGN--VGIVMSLEQIK-GIEDYMLLAGHLAMFTNDYVNLADLYLASSCPI 787
DB 661 TARKAFIRYQDLYRYLELISIEERKKGETYNNDLFLADVSYOGKFEHAALKYRSQSHEN 720
QY 788 AALEMRDLOHMSALQALAKHLARPDQIRFISKEVATOLEFAGQYVNLALAYEKGITGDNK 847
DB 721 LALEMTYDL-----CNFEYAKDFLG-----GSDPK 745
QY 848 EHEDEACLAGVAQMSIRMGDIRGVNQALKHPRSVLKRDGCAILENMKQFSEAAQLEYEKL 907
DB 746 E-----TKMLITKQADMAR-----NIKEPKAAVEMYSAG 775
QY 908 YTRKASVYIRSKNNAKVGDLPHVSSPKIHLQYAKKEDGKRYKEAVAYENAKOMOSV 967
DB 776 EHVKAIEI-IGDHWV--DMLIDI-----ARKIDKAREBPL-----L 810
QY 968 IRIYDHLNPEKAVINIVRETQSLODCAKKNVAFFLQDGYGSAIOFLVMSKCNNEAFPTLA 1027
DB 811 CATYLLKLDSPGYA-----AETYLKAGDLSLVQLHETQVDEDEAFALG 854
QY 1028 QOHNMKEIYADIIGSEDTNEDYOSIALYPEGKRYLQAGKFFLLCGOYRSLAHPLKCP 1087
DB 855 EKHPERK-----DDIYMPYQOMLAENDREFEQAARHAKNGROREAV----- 895
QY 1088 SSEDNVAIEMAIETVGOAKDELTLNOLIDHLLGENDGMPKDAKYLRRLYALKQYREAAQ 1147
DB 896 ----QVLEQTLNNNAVAESRF-----NDAAYYY--WMLSMQCDIDIAQ 930
QY 1148 TAITIAEEOGAGNYRNAHDVLFMSVAELKSQIKIPSEKATNMLILHSYILVKIHYKNG 1207
DB 931 DP---NOKDMLGKFH-----FORLAEL-----YHGTHAIHRTIEDP 955
QY 1208 DHMKGAMLRVANNISKFPSPHIVPILSTFVIECHRAGL---KNSAFSFAAMLMPREYR 1263

```

```

Db 966 FSVHRETLF-----NISRELFHSLPKDPTSGISKVRILFTLAKOSKALGAYRLAHAYD 1021
QY 1264 S-----KIDAKYKKIE---GMRRPDISEIEEATTPCPCKELLPECCCELLCPCCKRSITV 1316
Db 1022 KLRGLYTPAFQKISLGLTLTTAKRPFHSE-----LVPLC-----Y 1059
QY 1317 CIATGRHMLKDMWTCPHCFPALYS-----ELKIMLNTSTGCP 1355
Db 1060 RGSINPILNINIGNCINCRQPIFASADYVLHVEFYLEBGTDEALISLIDEVLMP 1119
QY 1356 MCSERLNAQOLKKSIDCTOYLRTVE 1380
Db 1120 KRDR-----OLEIANNSSQILRLVE 1140

RESULT 9
Q22830
ID Q22830 PRELIMINARY; PRT; 1758 AA.
AC Q22830;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE T27B1.1 PROTEIN.
GN T27B1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Cratton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spoat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RT Wu X., Le T.;
RT "The sequence of C. elegans cosmid T27B1."
RN Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
RL [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
CC -1. SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL; U41020; AA82332.1; -.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00515; TPR; 2.
DR Pfam; PF00400; WD40; 4.
DR SMART; SM00320; WD40; 3.
DR Repeat; WD repeat.
KW SEQUENCE.
SQ SEQUENCE 1758 AA; 198406 MW; 894255C3FEF7B0A6 CRC64;

```

Query Match 3.7%; Score 270.5; DB 5; Length 1758;
 Best Local Similarity 18.7%; Pred. No. 2.3e-09;
 Matches 270; Conservative 220; Mismatches 493; Indels 463; Gaps 68;

```

QY 65 SCNYLAVTADYIVKIFDRHGRKSEINLPNCVAMDMDDGDLVLAIEKSSCIYLMDA 124
Db 27 NSRAAIACSDSNVALDENGVOKDRF-----TCKPID-----AKYGRKSTTVLC--- 71

```

```

QY 125 NTNKTSQLDNGNRDQMSFLMSKVSFLAVGTGKGLXLYNHQTS--RKIPYLGKHTK-- 180
Db 72 -----TFSPDSRIALIGOSDNVLFYKVTGTSWNEKKVLYNKFQPS 113
QY 181 RTTCGGMANENLXALGGEKMTTYSNOEGDTTRQOVREBPANMOFFLMKMDRISAES 240
Db 114 AVTCLSWPDDKILVGQDLGKVRIG-----LITKNCSS-----LYKTD-----ET 154
QY 241 MISVVLGKRTLEFLNINEDPNADLEFOODFENIYCYNNYGGRIKIGESC----- 291
Db 155 VVSIOHPRKRTSFVSAH-----OD-GSIIILYNNSSRTQSKI--CLQVPPYVL 199
QY 292 ---GHEVVIS-----THGTGEGEIEQARNNKDNLTJASQTLNKV--ATCG--- 334
Db 200 VFTNGLVATSDRVLSTYENGAVVOQCPDYNDQSEKESISSCPAQNVAVTGCGVF 259
QY 335 ---DNCK-----IDQVLD-----LKDQVY-- 351
Db 260 SVDCCLRGMLKSRRETTYVAPSHVILRDVNTDRTNITSNGLAIDELKIMGKRRYVIG 319
QY 352 -----ILNDEENKGLGLTSMWDDGQLLALSTORGLHVFLLKPLIGDASTRIAYLTS 406
Db 320 YTSSTIITDTSQRRSELEW-----QSGHEKFTPD---FNCC-----LIIN 360
QY 407 LLEVTYAN--PVEGELPITVSVDPENFAVGL--YHLAVGNMNRAMFYVIGENAVKILK 462
Db 361 AGEVTFVEYGVDSIGWVTELTSPHLLSVQLKAFKSKYGRNFDPS-----ESSVSYSE 414
QY 463 DMEYLGTVASICLHSDYAAALFEGKVQHLISELIDQBEETRLFPRAVDKCRH---- 518
Db 415 HSDL--NTGQICVFSDFL-----QRTTSVFELIKAKKRIENDL 451
QY 519 -LCHALTSFLIYGTDTGVQVYFIEDMOFVNDYRHPVSVKRIFFDPNGTRLVFIDEKSD 577
Db 452 QLPHWIAPNLINGOQESINNTGAVDWIELNE-----RASKLYLNDKRS- 496
QY 578 GRYVCPVNDATETIPDFSTIGVL-----WENPMKQVFAIYDDKYYTVFHKD--- 629
Db 497 -----KVLVDISSDQSFVLSFCTYVQVPMSPDVIAQSGDNLSIV--YNDPJE 545
QY 630 -----TIQG-AKVIIAGSTKVPFAHKPRLLYNGELTCQOSQKV-----NIIYLSHGFLS 679
Db 546 QVTNKKIKGEVAVLADART-----EYVQPTAKVAYELDNTOIE---FGA 590
QY 680 NLKDXGPDELRLPMLANHMLKRFSDAMENCRIINDEAANMELARCLHMEVEFAIRYVR 739
Db 591 ALEKRDPRAVAFLESN--TSGTAYSM-----WIRVAMELEHGNLFVAQRCYA 638
QY 740 RIGNGVIVSLQIKGIED-----YNLAGHLAMFTNDYNLAQDIYIASSCP 787
Db 639 AINDYAKVRKLHDIIEIADIASISIGDGTHEYKVRAMLAINGRFRKEERIFELQNDTE 698
QY 788 AALEMRDLOHNDLSALQKLKHLAPDQIPISKEY-----AIOLEFA-GDYVNA 834
Db 699 SAIGWTSLHKWDEALELKVLTNPPEYEQKTSYLRALSDTQDSKAALAKYSDGDTLSA 758
QY 835 LAHYEK-----GITGDNKE-----HDEACIAGVAQMSIRMGDIRQVNOALHPSRVILK 884
Db 759 IOLYIKSNKPLSALSANNDVSLQDENILRQIADSLV-----SQLYD 802
QY 885 DCGAILENMKQFSEAQOLYEKLYYDKA-----ASVYTRSKNMKAVGDLRLHYVSP 935
Db 803 KAGDYVERKLKDEPKDAVEYFKGDAVGKALQARFAPEPVVLTLEQEWG----- 850
QY 936 KIHLOYAKAKEDGRY-----KEAVVAYENAKQWOSVIRIYLDHLNPEKAVNIVRET 988
Db 851 -IHLEYIGDYDAVNHVYVANDLKRKAVEAIPAKEN-----PALSLV--- 892
QY 989 OSIDGAKM-----VAREFLQDGYGSAIQFLYSKNCNNEAFTLAQOHNK---MEIYA 1037
Db 893 ENIQDQKVRTGYGELADHYSKNGDEPERARLFVEAGLGNDAIMYGNKNKXIDAFLRSE 952
QY 1038 DIIGSEDTTNEDYOSTALYFEGEKRYLQAGKFFLLCGOYSRLAKLHFLKCPSSSEDNVAIEM 1097

```

```

Db 953 EFHGREATIS--SYLAKAEDDEDEGRFAEAEOLYITIGMPHKAIOY----- 997
QY 1098 AIEYVOAKDELTLNLIHLLGENDGMPKDAYLFRLYALQYREAAOTAIILIREQ 1157
Db 998 --DRVGDDVVL---RLVERXGHEH--MHETRRF-----ATQYBERGD--LKAEEQ 1041
QY 1158 --SAGVNRNADVLFSMY--AELKSOKIKI-PSEMATNL--MILHSY-----ILVKIH 1203
Db 1042 FLAAGGFRSA-----VMNYKSEMSDAVRAKTEGGENMKOYLFPMKASIGGDAVKLL 1097
QY 1204 VKAGDHMKG-----ARMLIRVANNISKPPSHVPIILTSVITECHRAGLKNSAF 1251
Db 1098 NKHGMLMEGIDFACETGAFDLAFDLARIG---AKDRMGTVHVRLATQLE-EEGRLEDASK 1153
QY 1252 SFAMLMRPY-----RSKIDAYKKKIEGMVRRRPDISETTEATPTCPFC 1296
Db 1154 HYVEATKLNTYNTITWCOAVPSRFDLOKRLGNRPDLAVEMFIRDNMDADERVAK--DHC 1211
QY 1297 KFLPE 1302
Db 1212 ESLLPD 1217

RESULT 10
Q9HBG5 PRELIMINARY: PRT: 1292 AA.
AC Q9HBG5:
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE MDR10P-L.
GN MDR10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21134653; PubMed=11242542;
RA Gross C., De Baere E., Lo A., Chang W., Messiaen L.;
RT Cloning and characterization of human MDR10, a novel gene located at
RT 3q21 encoding a WD-repeat protein that is highly expressed in
RT pituitary and testis.;
RL DNA Cell Biol. 20:41-52(2001).
CC 1- SIMILARTY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AF244931; AAG15428.1; -.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 6.
DR SMART: SM00320; WD40; 7.
DR PROSITE: PS50082; WD_REPEATS_2; 1.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD Repeat.
SQ SEQUENCE 1292 AA; 147385 MW; 85F6493ED3B4A496 CRC64;

Query Match 3.7%; Score 268.5; DB 4; Length 1292;
Best Local Similarity 18.1%; Pred. No. 1.9e-09;
Matches 250; Conservative 203; Mismatches 512; Indels 413; Gaps 57;

```

```

Db 299 EERNDLIAVADMOQKVSFYOLSGKQIGKDBALNFDP-----CCISYFTGGEYILG---G 350
QY 293 HEVAVISTHNGELGOELFQARNHNDNLSTAVSOTLKKVATCGNCKIKIDVLKMYIYA 352
Db 351 SDKQVSLFT-----KDGV-----RLGTVG----- 369
QY 353 LNLDEENKGLTSLWDDGOLLASTQSGSLHVEFLKPLTGACSTRITAYLTSLEVIV 412
Db 370 ----EONSVMWTOGAPRDSYVVVGCGDGTISFYQLFSTVHGLYDRARLRDSMTDVIV 425
QY 413 ANPVEGELPTVSDVBPNEVAVGLYHLAVGMNRRAMFYVGENAVKKIKMEYLGTVAS 472
Db 426 QHLIT-EOKVRICKELVKIKAIYRRLAIQLPEKILIELYS---EDLSMHY----- 475
QY 473 ICHSDYMAALFEGKQVQLHLESEILDQEEER-ETLFLPRAVDK-----CRILCH 521
Db 476 --RVKERRIKKFECSNL-LVVCANHIIICQERLQCSFSGVKEREOMESLIRYIKVIG 532
QY 522 ALTSDFLIYCTDGTGVQYFYIEDMOQFVNDYRHRVSVKIFPPDNGTRLYFIDEKSDGFY 581
Db 533 PRREGILVGLKNGQILKIFVDMLFAVLILKQATAVRCIDMSRKRLKLVAVDENDCLVY 592
QY 582 CPVNDATYELIPDFSPTIKGYLVEMNMPDKGVT--AYDDKVTYVYVHKDPTIGARVILA 639
Db 593 ---DIDTKELLQEPNANSVAMNTQCEDMLCFSGGGYLNIKASTFPVHRQIKLG---FVV 646
QY 640 GSTKVPFAHNPRLLYNG-ELTCOTQSGKVNIIYLSHGFSLNKDKQPDRLRMLAHNL 698
Db 647 G-----YNGSKIFC-----LHVF--SISAVEPDSAPMYQY-LD 677
QY 699 LKRFSDAMEWCRIILNDEAANELARACLHMEVEFARVYRIGN---VGIVSLBOIK- 754
Db 678 RKLFEKAYQIACGVDTDMRELAMEALEGLDETAKKAFIRQDLRYLELISIERKK 757
QY 755 -GIEDYNLAGHLMFTNDYNLAQDLYLASSCPIALALEMRDLQHDSALQAKHLAPDQ 813
Db 738 QGFTNNDLFLADVFSYQGFKEHAALKYKRSGHENLALEMVTDL----- 780
QY 814 IPIFSKEVATQLEFADYVVALHKEGTTGDMKKEHDEACLAOVAOMSTFMGDIRGVNG 873
Db 781 -----CMFEYVADFLG-----SGDKE-----TKMLITQADMAR--- 810
QY 874 ALKHPKRLKRDGALLENKQFSEAAQOLYKEKLYUDKASVYIRSKNNAKVGDLLPHVS 933
Db 811 -----NIKEPKAAVEMYISGHEVKALEI-CGDHGW---DMLIDI- 847
QY 934 SPKIHLOYAKAKADGRYKAAVVAAYENAKOMQSVIRIYDLHNLNPEKAVNIVRETOSLDG 993
Db 848 -----ARKLDAEREPDL-----LCATYLLKLLDSPGYA----- 875
QY 994 AKVAVARFLOLGOYGAQLQYLVMSKCNNEAFTLAQCHNKKEIYADIIGSDTTNEDYOSI 1053
Db 876 ----AETYLKMGDLKSLVQHLHVEIQRWDEAFALGEKHPERK-----DDITMPY 919
QY 1054 ALYFEGEKRYLOAGKFFLLCGOYSRALKHLKCPSSSEDNVAIEMATYVGOAKDELTINO 1113
Db 920 AQMLAENDREFEAKQFHKRGQREAV-----QVLEQLLNNA 956
QY 1114 LIDLHLLGENDGMPKDAYLFRLYALQYREAAOTAIILAREQASAGVNRNADVLFSMY 1173
Db 957 VVESRF-----NDAAYY--WMLSMQCIDIADP---AQKDTMLKLFYH-----FORL 999
QY 1174 AELKSOKIKIRPSEMATNLMLHSYILVKKIVKNDHDKKAGMLIRVANNISKPPSHVPI 1233
Db 1000 AEL-----YHGHAHNRHTEDPSVNRPELTLE---NISRELLSLPK 1038
QY 1234 LTSTVIECHR---AGLKNASFPAAMLDPREYS---KIDAKYKKKIE---GMVRRPDI 1283
Db 1039 DTPSGISIKVAILFTLAKQSAALGAYRLARAHAYDKLRGLYIPAFQFSIELGTLIRAKPF 1098
QY 1284 SEIEEATTPCPFCFLLPBECCLCPGCKNSIPYCIATGRMLAKDMQVCPHCPPALYS- 1342

```

```

Db 1099 HNSEF-----LVPLC-----YRCSTNNPLNINCNVCINCROPIFSA 1136
Qy 1343 -----ELKIMINTESCPMCSERLNAOLKKISDCTOYLRTEE 1380
Db 1137 SSYDVHLVEFLIEBSITDEBAISLIDLEVLARKRDR-----QLEIANNSSQILRLVE 1190

RESULT 11
ID 09UFB0 PRELIMINARY: PRT: 1198 AA.
AC 09UFB0:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOHETICAL.137.1 KDA PROTEIN (FRAGMENT).
GN DKFZP434K016.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL; AL133565; CAB63718.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 5.
DR SMART; SM00320; WD40; 3.
DR PROSITE; PS50082; WD_REPEATS_2; 1.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
DR KW Hypothetical protein; Repeat; WD repeat.
FT NON_TER 1
SQ SEQUENCE 1198 AA; 137080 MW; EF3D17BB9B5DE88C CRC64;

Query Match 3.6%; Score 263; DB 4; Length 1198;
Best Local Similarity 18.2%; Pred. No. 4.1e-09;
Matches 245; Conservative 197; Mismatches 506; Indels 396; Gaps 54;

Qy 116 SSCIVLMDANTKTSOLDNMDQMSFLMSKVSFLAVGTAKGNLXIYNHQTSPRPV 174
Db 70 SSDFGIMSEPOKSVSK--HKSSKIIICSWINDGOTLALGMFNGIISLNKNGEEVKIE 127
Qy 175 -LGHKTRITCGCW-----NAENLXALGGEKMI-----TVSNQCG 209
Db 128 RRGGSLSPIWISICWNPSRMSFMMNRNEDA---EDVIVNYIQLIPSTLSAVYSSQG 184
Qy 210 DTIRTOTVNSEPXNNQFLLMKDDRTSAESMISVV-LGKTLFF-LNLNPDNPADLEF 267
Db 185 SEAESEEPEDDSR-----DDNLEERNDLAVADMOKVSYFOLSGKQIGKDALNF 238
Qy 268 QODFGIVVYNNYCGGR-IMIGESGHFVVISHTHGELOELFOARNHNDLSTAVST 326
Db 239 DP-----CTISFTYGEYLLG---GSDKQVSLFT-----KQGV----- 269
Qy 327 LNKVATCGDNCTIKIOLVDLKDMYVLLNDEENKGLSTWTDGQLALSTQSGSLHVF 386
Db 270 --RLGTVG-----EQNSWMTQOAKPDSYVYVVGQDGIISFY 305
Qy 387 LTKLILGDACSTRIVAYLTSLEAVYANPVEGBLPITVSVDVEPNFVAGLVLAGMNN 446
Db 306 QLIFTVHGLVKDRYAYRDSMTDVIYOHLLIT-EOKVRIKCKELVKKIAYIRRLAIQLPE 364
Qy 447 RAMEFVILGNNAKKLKDMEYLGTVASICHSDYAALFEFGKQVHLIESEILDAQEER-E 505
Db 365 KLILIELTS--EDLSDMHY-----RVKKEIIKKFECNL-LVYCAHHIILCCEKRLQ 412
Qy 506 TRLPFAVDK-----CRILCHALTSPLIYGTDTGVQVQVYFIEMQFVNDYRHV 555
Db 413 CLISFGVGEREMQESLRIYIKVIGGPRREGILVGLKNGQILKIFVDLFAIVLLKQAT 472

```

```

Qy 556 SVKTIFFPDNGTRLVEIDERSDGFVYCPVNDATYEIPDESPTIKGVLMENMDKGVFL 614
Db 473 AVRCLDMSASRKKLAVVENDOTCLVY---DIDTKELLFQEPANSAVANTQCEDMLCSG 529
Qy 615 -AYDDDKVYTYFHKDTTIGAKVITLAGSTKVPFAKPLILYNG-ELTCOTQSGKANNIYL 672
Db 530 GGYLNKASTFPVHHOKLOG---FVVG-----YNGSKIFC----- 561
Qy 673 STHGFLSNLKDGPDELRLAHNLKRFSDAMENCRILNDEAMNELARACLHMEVE 732
Db 562 -LHVF--SISAVEVQSAAMYQY-LDKILFKAYQIACLGIVDTMRRELAMALGLDPE 617
Qy 733 FAIRVYRIRGN--VGIWVSLQIK-GIEDYNLAGHLMFTNDYNLAQDLIYASSCPI 787
Db 618 TAKKAFIRVQDLRYELISSIERKKRGFTNNDLELADVFSYQGFHEAKIKRSGHEN 677
Qy 788 AALENRDLQHHDSALQAKHLAPPOIPISFEVIAQLEFADYNNALAHYKGTIGDNK 847
Db 678 LALEWYTDL-----CMFEYADFGL-----SGDPK 702
Qy 848 EHDEACLAGVAQMSIRMGDIRGVNQAALHPRSVLKRDGAILLENKQFSEAAQLYKGL 907
Db 703 E-----TKMLITRKQADMAR-----NIKEKKAVENTYISAG 732
Qy 908 YDKAASVYIRSKNAKVGDDLPHVSSPKIHQYAKAKADGRYKDAVYAVENAKOMQSV 967
Db 733 EHVAKIEI-CGDHGW--DMLIDI-----ARKLDAEREPEPL-----L 767
Qy 968 IRIYIDLNNPKAVNIVARETQSLDGAKMVAREFLOLGDYGAIDQFVMSKCNNEAFILA 1027
Db 766 CATYLLKLDSPGYA-----AETYLKMGDLKLVOLJHETQWMDAFAALG 811
Qy 1028 QOHNMEIYADIGSEDITNEDYOSIALYFESEKRYLAGKFFLLCGQYSRALKHFLKCP 1087
Db 812 EKHPEK-----DILYMPYAQWLAEENDPBEAKAFHKAAGQREAV----- 852
Qy 1088 SSEDNVAIEMAIETYQAKDELTLNQLIDHLLGENDGMPKDAKYFLRYLMALKOTREAAQ 1147
Db 853 -----QVLEQLTNNNAVRESR-----NDAAVYV--WMLSMOCLIDIAQ 887
Qy 1148 TAITAREEOSAGNRYNADVLFESWVAELKSQIKIRPSEMATNMLILSHYLVKIHVNG 1207
Db 888 DP---AQKDTMLGKTYH-----FQRLAEL-----YGHYAIHHTEDP 922
Qy 1208 DHMKGARMLIRVANNISKPPSHIVPILSTVIECHR---AGLKSASFSAAMLBPERS 1264
Db 923 FSVHREPTELF-----NISRLHSLPKDRPSSGISVKILFTLAKQSKALGAYRLARHAYDK 978
Qy 1265 ---KIDAKYKKKIE---GMVRPDISIEIEATTPCPFCCKFLLPCECLLPCKKSIYFC 1317
Db 979 LRGLYIPARFQKSIDELGTLITIRAKPFHDSSE-----LVPLC-----YR 1016
Qy 1318 IATGHHMLKDDMTVCPPHCFPALYS-----ELKIMINTESCPM 1356
Db 1017 CSTNPNPLNINCNVCINCROPIFSAASYDVHLVEFLIEBSITDEBAISLIDLEVLARK 1076
Qy 1357 CSERLNAOLKKISDCTOYLRTEE 1380
Db 1077 RDDR-----QLEIANNSSQILRLVE 1096

RESULT 12
ID 09HAT9 PRELIMINARY: PRT: 1241 AA.
AC 09HAT9:
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE SPG PROTEIN.
GN SPG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shan Y.X., Li J.M., Sha J.H.;
 RT "The research of spermatogenesis related genes."
 RL Submitted (Sep-2000) to the EMBL/GenBank/DBJ databases.
 CC - SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AF302154; AAG13415.1;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 6.
 DR SMART: SM00320; WD40; 7.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KM Repeat; WD repeat.
 SO SEQUENCE 1241 AA; 141824 MW; 6C3C543369A6BDF5 CRC64;

Query Match 3.6%; Score 263; DB 4; Length 1241;
 Best Local Similarity 18.2%; Pred. No. 4.3e-09;
 Matches 245; Conservative 197; Mismatches 506; Indels 396; Gaps 54;

QY 116 SSCIIYMDATNKTSDLNCGMRDMSFLMSKVSFLAVTGKGLXIIYHQTSPKIPV- 174
 DB 113 SSDFGLMSPPKSVSK--HKSSSKIIICSWTNDGQYLALGMPNGIISIRKNGEEKVIE 170
 QY 175 -LCKHTKRTICGW-----NAENLXALGEDKMI-----TVSNQEG 209
 DB 171 RRGSLSPISISICNPSRKESFMNRENDA--EDVIYNTYQETPSTLKSAYVSSOG 227
 QY 210 DTRQGVQVSEPNMOMFLMKMDRTSAESMISV-LGKRTLEF-LNLNPNPADLEF 267
 DB 228 SEAESEEPEDSDSPR-----DNLERNDILAVADMGGKSVFYOLGKQIGKDRALNF 281
 QY 268 QDQCGNIVCVNMGDGR-IMGFSCGHFVVISHTGSGEIGFQARHKNKNTLSIANST 326
 DB 282 DP-----CCISYFTKGYILLG--GSDKOVLSFT-----KQGV----- 312
 QY 327 LNKVATCGDMCIKIQDVLVDKDMYVILNDEENKGLTSLWTDGQQLALSTORGLHVF 386
 DB 313 --RLGTG-----EQNSWMTQCAKPRSNVVGCGCGGTISFY 348
 QY 387 LTKLPILGACSTRIVATLSLELVYANPYEGELPTVSVDEBNFVAVLHVLAVGMN 446
 DB 349 QLIFSTVHGILYKDRYAYRDSMTDVIYQHLIT-EDQVKRIKCKELVKAIAIYRNRLAQLPE 407
 QY 447 RANFYVGENAVKMKMEVLGTASTCLSDYAAALFEKGVQHLLESELDAQER-E 505
 DB 408 KILYELYS--EDLSDMY-----RVKEKIKKFCNL-LVGCANHILLCQEKRLQ 455
 QY 506 TRLEPAVDK-----CRILCHALTSDFLIYGTDTGVVQYFYEDMOFVNDYHRPV 555
 DB 456 CLSSSGKREKEMOMESLIRIKYIGGPRGREGVLGKNGOILKIFVNDLFAIYLKQAT 515
 QY 556 SVKKIFPDNGRLVFIDEKSDGEVYCPVNDATYEIPDSPTIKGVLMWMPMDKGVFI- 614
 DB 516 AVRCLDSASRRKLAAYVNDENTCLVY--DIDTKELLFQEPNANSVAMNTQCEMDLCEFSG 572
 QY 615 -AYDDKVVYVYVPHKDTIQCAKAVTLASTKVPFAHKRLLLXNG-ELTCQOSGQVNNIYL 672
 DB 573 GGYINIKASTFFVHRQLQG--FVVG-----YNSKRIK----- 604
 QY 673 STHGFLNLKDXGDEPLMLANLMLKRFSDAMEMCRILNDELAAMELRACLHHEVE 732
 DB 605 -LHVF--SISAVEPQASAPMYQ-LDRKRLKFAVQIACLGVTDTDMELAMEALEGLDFE 660
 QY 733 PAIRVYRIGN--VGIVMSLEQIK--GIEDYNLLAGHLAMFTNDYMLADLYLASSCPI 787
 DB 661 TAKKAFIRVQDLRYLELISIEERKKRGETNNDLFLADVESYQCKFHEAALYKRSCHEN 720
 QY 788 AALEMRDLDQWDSALQAKHLADQIPFISKVATIQLEFADGVNMLAHNEKITTDGK 847
 DB 721 LALEMVYDL-----CMFEYAKDFLG-----SDDPK 745

QY 848 EHDACIAGVAGWMSIRMGDIRGVNQAALKHPSHVLRKDCGALLENNKQFSEAQLYEKGL 907
 DB 746 E-----TKMLITKQADMAR-----NIKEPAAYEMVTSAG 775
 QY 908 YYDKASVYIRSKNMKAVGCLLPVSSPKTHLOAKAKEDGKREAVVAYENAKOMQSV 967
 DB 776 EHVKAIEI-CGDHGW--DMLIDI-----ARKLDKKEERFLL-----L 810
 QY 968 IRIYDLNPNPEKAVNIVRTQSLDGKAVARFLOLGDGSGAIOPFLVMSKNNEAFTLA 1027
 DB 811 CATYLLKLDSPGYA-----AETYLKMGDLKSLVQLHETVORNDENAFALG 824
 QY 1028 QOHNMKEIVADIIGSEDTTNEYOSIALYPEGEEKRYLQAGKFFLLCGOYSRALKHFLCP 1087
 DB 855 EKHPEPK-----DDIIMPYQWLAENDRFEQAARHKHKGROREAV----- 895
 QY 1088 SSEDNVAIEAIEITGQAKDELLTNQIDHLLGBNDGMPYDAKYLFRLYALKOYREAQ 1147
 DB 896 -----QVLEQLTNNAVAESRF-----NDAAYVY--WMLSMQCLDIAQ 930
 QY 1148 TAIITAEBSAGYRANAHVLFMSVYELKSQKIKIPSEMATNMLHSYILVKIKHNG 1207
 DB 931 DP--AOKDMLGFEYH-----FORLAEL-----YHGTHAIHRTEDP 965
 QY 1208 DHMKGARMLIRVANNISKPSHIVPILSTVIECHR--AGLKNASFFSFAMLMRPEYRS 1264
 DB 966 FSVHREPTEL-----NISRFLLHSIPKPTPGISGVKILFLFLAQSKALGAVRLARHAYDK 1021
 QY 1265 ---KIDAKKKKIE--GNVRRPDISEIEEATTPCPCKFLLEPCELLCPGCKNSIPYC 1317
 DB 1022 LRGLYIPARFOKSIETLTLIRAKPFHDSB-----LVPLC-----YR 1059
 QY 1318 IATGRHMLKDMWYCPHCPDPALYS-----ELKIMLTNESTCPM 1356
 DB 1060 CSTNNPLNLDGNCVCIKCRPFIFSASSYVHLVRYLEEGITDEPAISLIDLEVLRPK 1119
 QY 1357 CSERLNAQLKISDCTQYLRTEE 1380
 DB 1120 RDR-----QLEIANNSQILRLIVE 1139

RESULT 13
 060332
 ID 060332 PRELIMINARY; PRT; 1462 AA.
 AC 060332;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE KIAA0590 PROTEIN.
 GN KIAA0590.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
 RA Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 5:31-39(1998).
 DR EMBL: AB011162; BAA2516.1; -;
 DR InterPro: IPR002885; PPR.
 DR SMART: SM00320; WD40; 1.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KM Repeat; WD repeat.
 SO SEQUENCE 1462 AA; 165196 MW; B05CF4325F364F0C CRC64;

Query Match 3.5%; Score 256.5; DB 4; Length 1462;
 Best Local Similarity 17.3%; Pred. No. 1.6e-08;
 Matches 247; Conservative 219; Mismatches 485; Indels 475; Gaps 64;

54 GAPIQAMOKTSGNVLAV-----TGADYIVKIFDRHGO-KRSLINLPKNCVADMDK 104
 16 GSPFSLMHPVH-PFLAVAVISTTSGS--VDILEGECVPDTHVERFRVAVSLCMPH 71
 105 DGDVLAVIAEKSSCIYIMPANT-----NKTSQLDNGM----RDQMSFLMSKVGSFLAVGT 156
 72 TRVLAV-----GWEIYEVTENKODKECHTMPRLHTADITVLRSPSNCLISGD 122
 157 VGNLKIYHQRSKI---PVL---GKHTKRITGCGW-----NAENL-----XALGED 199
 123 RGLVLLMLRLDQGRVQGRPRLKHEYGKHLTH---CIFRLPPGSDILVOLAKAASGDE 178
 200 KMTVSNQSGDTIRQOVASEPANKMOFLMKD-----DRISAASMTISVVLGKTLTF 253
 179 KALDMFNMKSSSGSLIKMGSHGELLEFSVLMGTVHYVDEKGTQVVASDSTIOMLFY 238
 254 LLINE-----PDNPAD-----LEFOQDFGNIVCYMWYDGRIMIGFSG 292
 239 MEKRELVVYVTEMLSLITVTPREGKAEEVMKYLSGKTR-----RADIALIG 288
 293 HEVYISTHFG-----ELGQETFOQRNHK-----DNLTSAVSQTLNKVATCGDN--- 336
 289 SLIVMAVGAALRFMDIERGENYILSPDEKFEKENMKVCYCKVKGHLAAGTRGRV 348
 337 --CIRKQDLY-----DLKDMYVILINDENKKGITLSTWDDQGLLALSTQSGSLHFLRK 389
 349 AMRKVPDLGSGAGCKRMALQPTLEOGNTTQIQMSRKMLLVANS-----VIS 400
 390 LPLIAD-ACSTRIAYVLSLEYVANPVEGELPIYVSVDEPFAVAVGH----- 439
 401 VALLSRAAMSHHQVAAQVVS-----PSLLNV-----CFELSTGAHSLRTMHIS 447
 440 -----LAVGNMNRAMTYVJGANA VKLKDMEYLTGVAASICLSHSDYAAALFEGVQLH 491
 448 GFVATKDAVAVMNRQVAFELSGAIIRSAG--TFLCETPVLAMHDEENYTVESNVOVR 505
 492 LIESEILDQOEERETRLPFAVDDKCRILCHALTSIDLTIGTGVQVQYFIEDMOWVNDY 551
 506 TWQGYVKOLLSETSEIGNCFELDIC-----GNFLVGTDLALFKSDLSRR- AKH 556
 552 RHPVSVKILP-----DENGTRLVFIDEKSDG-----FVYCPVNDATYEIPDFSP 596
 557 CCSRSLAEIVPGVGIASLRCSGSGSTISILPSKADNSPDSKICFYDVMDDITVTFDE-- 614
 597 TTKGVLMEMMPMDKGVFIAYDDKVITYVFHKDTIOGAKVILAGSTKVFPAH-----KFL 651
 615 -----KTGOIDRRETLSEFNEOE--TNKSHLFEVDEGLK-----NYVUNHFWDOSEBR 659
 652 LLY-----NGELTQOTQSGK-----VNNIYLS--THEFLSLMKXGPD 687
 660 LVCVCAVQETPSPQOSANG-----QPQDGRAGPADVILISFFISEHGF--LHESFP- 712
 688 ELRPMLANLM-----SDANEMCKIILNDEAAMFLAACLHMEVEPAIRYRIGAV 744
 713 --RPATSHSLTMEVYUYFTTRKPEADREDEVEPCHHIIPQWVNSRPLRDFGLEDCK 770
 703 -----SDANEMCKIILNDEAAMFLAACLHMEVEPAIRYRIGAV 744
 771 ATRDAMLHSEFVTITGDMDEAFKSIKLKSEAVWEMMAMCVKTORLDAVA-RVC--LGMN 827
 745 GIVMSLEQKJIEDNVLNLAGHLMFTNDYNLAODL-YLASSCPIALALEMRRLQH----- 798
 828 GARARARALREKEQEPLEEARVAVLATOIGMLEDAEQILTKC-----KRHDLNLFYQA 881
 799 ---WDSALQALAKHLAPDQIPFTISKEXAYLOLEFAGDYVNALAHYEKGTGDNKEHDEACIA 855
 882 AGRMQEALQVAAHHDRVHLRSTYHRYAGHLEASADCSRALSLEYEKSDT----- 929

856 GVAQMSIRMGDIRGVNQA LKHPSRVLRKDCGAILIENMKOFSEAAOLYERGLYIDKAASV 915
 930 -----HRFEVP-RMLSEDLPSL-----ELYVK----- 951
 916 YIRSKN-----MAKVGDLLPHVSSPKTHILOAKAKREADGRYKEAVYENAKOMOSVIRY 971
 952 -MKDKTLMRWMA-----OYL---ESQGMDAALAHYELARHFSLSVRTH 991
 972 LDHLNPKRAVNIYREFTQSLDCAKNVAREFLOLGDGSAIOFLVMSCKNNEAFTLAAQHN 1031
 992 C-FQGVQVQAQAOANTGTGLASVYHARQYEOOEYGAOVHHYTRQAQAKNIRLCKEN- 1049
 1032 KMEITADITIGSED-----TTNEDYOSIALYFBESEKRYLOAKGPELL---CGQTSRAL 1080
 1050 -----GLDDQMLNALLLSSPEDEMIIVARYY--EEKGVOMDRAVMLYHRAGHFSKAL 1098
 1081 KHFLEKPSSEDNVAIEMALEYVGOAKDELITNOLDHLLGENDMPKDAKY--LRLVYA 1138
 1099 ELAF---ATQGFVALQDLIEDETSDPALARCSDFI-----EHSQYRAVELLIA 1148
 1139 LKQYREAOATAIIRAEOSAGNYRNAHDVLFMSYAEKLSQRIKIPSEMATIMILHS-- 1196
 1149 ARKYOALQGL-----GQMSITEEMAEKMTYAKDSS 1181
 1197 -----YLVKIHVKNQDMMKGMRLIR 1218
 1182 DLPEESRRELLQIADCCMRQSGSYHLATKKYQAGNKLKAMRALIK 1227

RESULT 14
 090G01 PRELIMINARY; PRT; 1653 AA.
 AC 090G01;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE HYPOTHEICAL 186.7 KDA PROTEIN (FRAGMENT).
 GN DKEZP434A163.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN NCBI_TaxId=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Wambuit R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL110218; CAB53678.1; -
 KW Hypothetical protein.
 FT NON-TER
 SQ SEQUENCE 1653 AA; 186734 MW; 670628A36CA762D4 CRC64;

Query Match 3.5%; Score 251.5; DB 4; Length 1653;
 Best Local Similarity 19.0%; Pred. No. 4.3e-08;
 Matches 243; Conservative 181; Mismatches 461; Indels 391; Gaps 58;

179 TKRITCGMNAENLXALG---GEDKMITVSNQSGDTIRQYVSEPNXNQFLMKMDRT 235
 16 TSAVVCLOMPAYIIVFGLABGVRLANKTKTKSSITGT-----ESVVSLLTINC 66
 236 SAAESNISVVLGKTLFLPLNLNEPDPNADLEFQDFGNIVC---YNNYDGRVIGFSG 291
 67 SGKGLISGHADTIVRYFF---DDEGSGESQGLVNHPCPPYALAVATNSIVAAG--C 119
 292 GHFVYISTHTGSGEIQARHKNLNTLSIANSQTLNKVATCGD-NCIKIQDVLVCLKMY 350
 120 DKRIYAYGKEGHMLQTFDYSRDQREFFTAVSSPGGGSVILGSYDRLLVFMWIPRRSIW 179
 351 VILNDEENK-----GLGTLSTWDGQLALSTQSGSLHVFITKPLIADACSTRIA 402
 180 -----EAKKPEITNLTITITALAKRDCGSRICVGTLCGGVGF-----DCLLRSI 225

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 14:50:35 ; Search time 38.73 Seconds
(without alignments)
3080.125 Million cell updates/sec

Title: US-09-729-653-2_COPY_1_1074
Perfect score: 1074
Sequence: 1 HSLIGRCSRGILGDNVAC.....LYFGEKRYLQAKFFLLCG 1074

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_032802.*

```
1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	367	34.2	502	22	AAW79303	Human protein SEQ
2	277	25.8	515	22	AAM80287	Human protein SEQ
3	80	7.4	138	22	AAU14751	Novel bone marrow
4	62	5.8	244	22	AAU14657	Novel human secret
5	32	3.0	72	22	AAU33013	Novel human secret
6	25	2.3	237	22	AAU33013	Novel human secret
7	19	1.8	109	22	AAU33123	Novel human secret
8	7	0.7	55	20	AAU41351	Human secreted pro
9	7	0.7	57	14	AAR38501	M. bovis pilin pro
10	7	0.7	58	22	AAU48729	Mouse liver growth
11	7	0.7	61	21	AAU03014	Human secreted pro

12	7	0.7	62	21	AAB39306	Human secreted pro
13	7	0.7	63	22	ABB44131	Peptide #11637 enc
14	7	0.7	63	22	ABB17718	Human nervous syst
15	7	0.7	63	22	ABB27013	Protein #9012 enco
16	7	0.7	63	22	AAW65154	Human bone marrow
17	7	0.7	63	22	AAW77859	Human bone marrow
18	7	0.7	63	22	AAU21759	Peptide #8193 enco
19	7	0.7	63	22	AAU38081	Peptide #12118 enc
20	7	0.7	65	21	AAB58410	Lung cancer associ
21	7	0.7	65	22	AAU44940	Protonibacterium
22	7	0.7	77	22	AAU57398	Protonibacterium
23	7	0.7	85	22	AAU76182	Human colon cancer
24	7	0.7	86	21	AAU26901	Zea mays protein f
25	7	0.7	88	22	AAU89209	Human immune/haema
26	7	0.7	90	22	AAU33124	Novel human secret
27	7	0.7	98	22	AAU47185	Protonibacterium
28	7	0.7	98	22	AAU91472	Human immune/haema
29	7	0.7	106	22	AAU34835	Chlamydia pneumoni
30	7	0.7	110	21	AAU28487	Zea mays protein f
31	7	0.7	114	20	AAU74083	Human prostate tum
32	7	0.7	114	20	AAU01147	Secreted protein e
33	7	0.7	117	21	AAU22616	Zea mays protein f
34	7	0.7	119	21	AAU64673	Human 5' EST relat
35	7	0.7	130	21	AAU43607	Arabidopsis thalia
36	7	0.7	135	21	AAU43606	Arabidopsis thalia
37	7	0.7	137	21	AAU26900	Zea mays protein f
38	7	0.7	140	21	AAU55757	Arabidopsis thalia
39	7	0.7	140	21	AAU61480	Arabidopsis thalia
40	7	0.7	140	22	AAU65756	Testis Enhanced Ge
41	7	0.7	141	22	AAU68476	Drosophila melanog
42	7	0.7	144	21	AAU15956	E. coli proliferat
43	7	0.7	144	22	AAU34747	E. coli cellular p
44	7	0.7	144	22	AAU35523	Haemophilus influe
45	7	0.7	144	22	AAU38506	Salmonella typhi c
46	7	0.7	144	22	AAU98241	Escherichia coli p
47	7	0.7	144	22	AAU98845	E. coli growth and
48	7	0.7	151	21	AAU27649	Secreted protein A
49	7	0.7	151	18	AAU44085	Human secreted pro
50	7	0.7	151	21	AAU10239	Human adult ovary
51	7	0.7	151	21	AAU13028	Arabidopsis thalia
52	7	0.7	151	21	AAU22615	Zea mays protein f
53	7	0.7	151	21	AAU25116	Arabidopsis thalia
54	7	0.7	153	21	AAU42871	Arabidopsis thalia
55	7	0.7	153	22	AAU68216	Drosophila melanog
56	7	0.7	156	22	AAU52660	Protonibacterium
57	7	0.7	161	21	AAU13027	Arabidopsis thalia
58	7	0.7	161	21	AAU25115	Arabidopsis thalia
59	7	0.7	161	21	AAU42870	Arabidopsis thalia
60	7	0.7	163	20	AAU31928	Vernonia 1,3-beta-
61	7	0.7	170	21	AAU43605	Arabidopsis thalia
62	7	0.7	203	18	AAU28268	Amino acid sequenc
63	7	0.7	229	21	AAU44501	Glutathione-S-tran
64	7	0.7	232	21	AAU13026	Arabidopsis thalia
65	7	0.7	232	21	AAU42869	Arabidopsis thalia
66	7	0.7	235	22	AAU52773	Protonibacterium
67	7	0.7	249	13	AAU20144	Class B beta-lacta
68	7	0.7	249	13	AAU20573	Class B beta-lacta
69	7	0.7	249	13	AAU20574	Class B beta-lacta
70	7	0.7	258	21	AAU19279	Arabidopsis thalia
71	7	0.7	258	21	AAU48498	Arabidopsis thalia
72	7	0.7	269	21	AAU22220	Arabidopsis thalia
73	7	0.7	272	22	AAU66438	Human ATPase 30.
74	7	0.7	276	21	AAU48497	Arabidopsis thalia
75	7	0.7	277	22	AAU15556	Novel human diapo
76	7	0.7	284	21	AAU39054	Arabidopsis thalia
77	7	0.7	293	21	AAU22219	Arabidopsis thalia
78	7	0.7	295	21	AAU30200	Arabidopsis thalia
79	7	0.7	296	20	AAU36855	Protein Involved i
80	7	0.7	296	21	AAU35215	Enterococcus faeca
81	7	0.7	306	21	AAU22218	Arabidopsis thalia
82	7	0.7	310	21	AAU48496	Arabidopsis thalia
83	7	0.7	312	22	AAU27719	Human full-length
84	7	0.7	315	19	AAU98227	H. pylori GHPo 115

85	7	0.7	316	22	AAU48530	Proionibacterium
86	7	0.7	322	22	ABBS5960	Drosophila melanog
87	7	0.7	328	21	AAI6118	Arabidopsis thalia
88	7	0.7	330	21	AAI39053	Arabidopsis thalia
89	7	0.7	334	21	AAI4646	Arabidopsis thalia
90	7	0.7	336	22	AAU27891	Human contig polyp
91	7	0.7	338	15	AAI65965	T. niveum GAPDH.
92	7	0.7	340	21	AAI75131	Neisseria meningit
93	7	0.7	341	22	ABG29988	Novel human diagno
94	7	0.7	344	21	AAI75130	Neisseria meningit
95	7	0.7	345	21	AAI30199	Arabidopsis thalia
96	7	0.7	345	22	AAI3858	Staphylococcus aur
97	7	0.7	356	22	AAU00858	S. aureus D-alalan
98	7	0.7	359	17	AAI88870	Sardinian tomato y
99	7	0.7	359	17	AAI88871	Sardinian tomato y
100	7	0.7	359	17	AAI88872	Sardinian tomato y
101	7	0.7	361	8	AAI70562	Product of ORF 4 f
102	7	0.7	362	15	AAI48694	G-protein coupled
103	7	0.7	362	17	AAI02666	G-protein coupled
104	7	0.7	364	21	AAI51230	Newcastle disease
105	7	0.7	373	20	AAI41692	Human PRO 363 prot
106	7	0.7	373	21	AAI44248	Human PRO363 (UNQ3
107	7	0.7	373	21	AAI33430	Human PRO363 prote
108	7	0.7	373	22	AAI12365	Human PRO363 polyp
109	7	0.7	373	22	AAI48108	Human A236 polypep
110	7	0.7	373	22	AAI48126	Mouse A236 polypep
111	7	0.7	373	22	AAI48145	Human A236 variant
112	7	0.7	373	22	AAI48146	Human A236 variant
113	7	0.7	373	22	AAI48147	Human A236 variant
114	7	0.7	373	22	AAI48148	Mouse A236 variant
115	7	0.7	373	22	AAI48149	Mouse A236 variant
116	7	0.7	373	22	AAI48150	Mouse A236 variant
117	7	0.7	373	22	AAI65293	Human PRO363 prote
118	7	0.7	373	22	AAI65293	Murine adipocytes-
119	7	0.7	379	22	AAI52464	Mycobacterium ture
120	7	0.7	393	22	AAI36802	Staphylococcus aur
121	7	0.7	403	21	AAI37514	Arabidopsis thalia
122	7	0.7	404	21	AAI06781	Arabidopsis thalia
123	7	0.7	404	21	AAI38369	Arabidopsis thalia
124	7	0.7	405	21	AAI37513	Arabidopsis thalia
125	7	0.7	410	22	AAI04102	Botulism toxin hea
126	7	0.7	411	21	AAI39052	Arabidopsis thalia
127	7	0.7	426	22	AAI3967	Human colon cancer
128	7	0.7	433	21	AAI06780	Arabidopsis thalia
129	7	0.7	433	21	AAI38368	Arabidopsis thalia
130	7	0.7	435	22	AAI95505	Human protein sequ
131	7	0.7	439	21	AAI06779	Arabidopsis thalia
132	7	0.7	439	21	AAI38367	Arabidopsis thalia
133	7	0.7	443	22	AAI14645	Arabidopsis thalia
134	7	0.7	443	22	AAI20005	Arabidopsis 3-keto
135	7	0.7	446	21	AAI14644	Arabidopsis thalia
136	7	0.7	446	22	ABG29986	Novel human diagno
137	7	0.7	447	21	AAI37512	Arabidopsis thalia
138	7	0.7	448	20	AAI03773	Human T-box polype
139	7	0.7	450	21	AAI69997	Human receptor-ass
140	7	0.7	456	21	AAI57017	Human prostate can
141	7	0.7	462	21	AAI30198	Arabidopsis thalia
142	7	0.7	462	22	AAI20004	Arabidopsis 3-keto
143	7	0.7	476	11	AAI05599	BIV gag gene produ
144	7	0.7	480	22	ABI64489	Drosophila melano
145	7	0.7	490	22	AAI20009	Brassica 3-ketoacy
146	7	0.7	510	20	AAI22201	Human extracellula
147	7	0.7	510	21	AAI54368	Human extracellula
148	7	0.7	510	21	AAI66646	Protein encoded by
149	7	0.7	510	22	AAI03653	Membrane-bound pro
150	7	0.7	510	22	AAI65169	Human extracellula
151	7	0.7	510	22	AAI50955	Human PRO698 (UNQ3
152	7	0.7	525	22	AAI04880	Human protease pro
153	7	0.7	525	22	AAI94023	Human protein sequ
154	7	0.7	546	22	ABG2395	Novel human diagno
155	7	0.7	566	20	AAI23888	Auifex VF-5 DNA p
156	7	0.7	574	20	AAI00940	Agf signal peptide
157	7	0.7	600	21	AAI44500	GST signal peptide
158	7	0.7	640	20	AAI7762	Maize lysine ketog
159	7	0.7	675	21	AAI10163	ISSI/1 polymerase
160	7	0.7	684	21	AAI0162	ISSI/1 polymerase
161	7	0.7	706	22	AAI79588	Corynebacterium g1
162	7	0.7	745	22	ABI56741	Drosophila melanog
163	7	0.7	750	21	AAI31281	Arabidopsis thalia
164	7	0.7	752	17	AAI97834	Kaposi's sarcoma a
165	7	0.7	752	17	AAI93610	Kaposi's sarcoma a
166	7	0.7	753	22	ABI67122	Drosophila melanog
167	7	0.7	760	21	AAI31280	Arabidopsis thalia
168	7	0.7	769	21	AAI31279	Arabidopsis thalia
169	7	0.7	777	22	ABI60046	Drosophila melanog
170	7	0.7	798	22	ABI62893	Drosophila melanog
171	7	0.7	799	21	AAI40275	Human ORFX ORF39 p
172	7	0.7	811	22	AAI31166	Novel human secret
173	7	0.7	818	22	ABI63426	Drosophila melanog
174	7	0.7	828	21	AAI44498	GST signal peptide
175	7	0.7	874	22	AAI46696	Fowlpox virus DNA
176	7	0.7	889	22	AAI68948	Arabidopsis thalia
177	7	0.7	906	22	AAI81162	Mycobacterium tube
178	7	0.7	1019	22	AAI92745	C glutaminc prote
179	7	0.7	1022	20	AAI87761	Maize lysine ketog
180	7	0.7	1091	21	AAI38615	Arabidopsis thalia
181	7	0.7	1096	21	AAI38614	Arabidopsis thalia
182	7	0.7	1139	22	ABG22397	Novel human diagno
183	7	0.7	1140	21	AAI38613	Arabidopsis thalia
184	7	0.7	1333	22	ABI66754	Drosophila melanog
185	7	0.7	1346	21	AAI39112	Arabidopsis thalia
186	7	0.7	1366	22	ABI65570	Drosophila melanog
187	7	0.7	1390	21	AAI39111	Arabidopsis thalia
188	7	0.7	1403	21	AAI39110	Arabidopsis thalia
189	7	0.7	1430	22	ABI58724	Drosophila melanog
190	7	0.7	1446	22	ABG29987	Novel human diagno
191	7	0.7	1455	22	AAI79120	Human protein SEQ
192	7	0.7	1612	19	AAI56088	R. prowazekii S-la
193	7	0.7	1621	22	AAI25646	Human protein sequ
194	7	0.7	1668	22	AAI34158	Staphylococcus aur
195	7	0.7	1703	22	ABI66223	Drosophila melanog
196	7	0.7	1953	22	ABI62633	Drosophila melanog
197	7	0.7	2058	22	AAI97070	Human polypeptide
198	7	0.7	2397	22	AAI6672	Staphylococcus aur
199	7	0.7	2435	22	ABI60448	Drosophila melanog
200	6	0.6	7	21	AAI19157	Peptide derived fr
201	6	0.6	9	21	AAI01463	Minor histocompati
202	6	0.6	10	18	AAI40685	Peptide which bind
203	6	0.6	10	18	AAI40691	Peptide which bind
204	6	0.6	10	18	AAI40698	Peptide which bind
205	6	0.6	10	20	AAI43535	Leishmania mexican
206	6	0.6	10	21	AAI01461	Minor histocompati
207	6	0.6	11	16	AAI79901	Human FK-506 cytos
208	6	0.6	13	20	AAI23261	Murine apolipoprot
209	6	0.6	14	19	AAI40274	K. oxytoca R-speci
210	6	0.6	14	20	AAI23254	Murine apolipoprot
211	6	0.6	14	20	AAI23256	Apolipoprotein E d
212	6	0.6	14	20	AAI23257	Apolipoprotein E d
213	6	0.6	14	20	AAI23258	Apolipoprotein E d
214	6	0.6	14	20	AAI23259	Apolipoprotein E d
215	6	0.6	14	21	AAI67877	Steap/Gbeta bindin
216	6	0.6	15	20	AAI23260	Murine apolipoprot
217	6	0.6	15	21	AAI38099	Caenorhbditis ele
218	6	0.6	15	22	AAI99455	Vaccine related MH
219	6	0.6	15	22	AAI66557	Human cytomagalovl
220	6	0.6	15	22	AAI66559	Human cytomagalovl
221	6	0.6	15	22	AAI66561	Human zinc finger
222	6	0.6	16	22	AAI99816	Vaccine related MH
223	6	0.6	17	16	AAI85128	Cholera toxin B an
224	6	0.6	17	22	AAI99553	Vaccine related MH
225	6	0.6	20	22	AAI12576	Human 5' EST secre
226	6	0.6	21	22	AAI89179	HIV gp120 protein
227	6	0.6	21	22	AAI89180	HIV gp120 protein
228	6	0.6	21	22	AAI89181	HIV gp120 protein
229	6	0.6	22	22	AAI47407	Peptide #37 for iL
230	6	0.6	23	15	AAI66107	PUR4421-encoded pe

231	6	0.6	24	21	AAV87538	Mature conotoxin p	304	6	0.6	53	22	AAU64777	Propionibacterium
232	6	0.6	25	10	AAV93093	CD4 anti-receptor	305	6	0.6	53	22	ABG09316	Novel human diagno
233	6	0.6	25	10	AAV09081	EGFP mutated fragm	306	6	0.6	54	20	AAV29016	T. gondii immunoge
234	6	0.6	32	19	AAV20979	Human gliat fibril	307	6	0.6	54	22	AAU60158	Propionibacterium
235	6	0.6	34	21	AAV38503	Fragment of human	308	6	0.6	54	22	AAU25487	Propionibacterium
236	6	0.6	35	22	ABV38022	Peptide #5528 enco	309	6	0.6	55	22	AAE01648	T. gondii immunoge
237	6	0.6	35	22	ABV32356	Protein #5255 enco	310	6	0.6	56	21	AAU03395	Human secreted pro
238	6	0.6	35	22	AAV58649	Human brain expres	311	6	0.6	56	22	AAE01649	Human gene 18 enco
239	6	0.6	35	22	AAV71155	Human bone marrow	312	6	0.6	57	22	ABG24412	Novel human diagno
240	6	0.6	35	22	AAV31436	Peptide #5473 enco	313	6	0.6	57	22	AAU03197	Novel human diagno
241	6	0.6	36	22	AAV89533	Human immune/haema	314	6	0.6	58	21	AAV51413	Human polypeptide
242	6	0.6	38	22	ABG16492	Novel human diagno	315	6	0.6	58	22	AAU43788	Human secreted pro
243	6	0.6	38	22	ABG19318	Novel human diagno	316	6	0.6	58	22	ABG28760	Propionibacterium
244	6	0.6	40	21	AAV16889	Bacteriophage Dp-1	317	6	0.6	58	22	ABV30120	Novel human diagno
245	6	0.6	40	22	ABV28824	Human peptide #875	318	6	0.6	58	22	ABV35292	Peptide #2771 enco
246	6	0.6	40	22	ABV33399	Peptide #905 enco	319	6	0.6	58	22	AAV96030	Peptide #2798 enco
247	6	0.6	40	22	ABV18858	Protein #857 enco	320	6	0.6	58	22	AAU18878	Human reproductive
248	6	0.6	40	22	AAV54184	Human brain expres	321	6	0.6	58	22	AAV56121	Novel prostate gla
249	6	0.6	40	22	AAV66578	Human bone marrow	322	6	0.6	58	22	AAV68493	Human bone marrow
250	6	0.6	40	22	AAV14451	Peptide #885 enco	323	6	0.6	58	22	AAV16301	Peptide #2735 enco
251	6	0.6	40	22	AAV26864	Peptide #801 enco	324	6	0.6	58	22	AAV04035	Peptide #2717 enco
252	6	0.6	40	22	AAV02178	Peptide #860 enco	325	6	0.6	59	22	AAV46459	Propionibacterium
253	6	0.6	41	21	AAV01464	Minor histocompati	326	6	0.6	59	22	ABG25720	Novel human diagno
254	6	0.6	41	21	AAV01465	Minor histocompati	327	6	0.6	60	22	AAU39131	Propionibacterium
255	6	0.6	41	21	AAV011979	Human polypeptide	328	6	0.6	60	22	AAV51250	Propionibacterium
256	6	0.6	42	22	ABV28150	Human peptide #801	329	6	0.6	60	22	ABG24749	Novel human diagno
257	6	0.6	42	22	ABV28167	Human peptide #818	330	6	0.6	60	22	ABV29102	Peptide #1753 enco
258	6	0.6	42	22	ABV30177	Peptide #2828 enco	331	6	0.6	60	22	ABV34265	Peptide #1771 enco
259	6	0.6	42	22	ABV33325	Peptide #831 enco	332	6	0.6	60	22	ABV19698	Protein #1697 enco
260	6	0.6	42	22	ABV33342	Peptide #848 enco	333	6	0.6	60	22	AAU21193	Human novel foetal
261	6	0.6	42	22	ABV35343	Peptide #2849 enco	334	6	0.6	60	22	AAV55056	Human brain expres
262	6	0.6	42	22	ABV18785	Protein #784 enco	335	6	0.6	60	22	AAV67448	Human bone marrow
263	6	0.6	42	22	ABV18802	Protein #801 enco	336	6	0.6	60	22	AAV15274	Peptide #1708 enco
264	6	0.6	42	22	ABV20784	Protein #2783 enco	337	6	0.6	60	22	AAV27740	Peptide #1771 enco
265	6	0.6	42	22	AAV54128	Human brain expres	338	6	0.6	60	22	AAV03019	Peptide #1701 enco
266	6	0.6	42	22	AAV56175	Human brain expres	339	6	0.6	61	21	AAV03138	Human secreted pro
267	6	0.6	42	22	AAV66504	Human bone marrow	340	6	0.6	61	22	AAV52545	Propionibacterium
268	6	0.6	42	22	AAV66521	Human bone marrow	341	6	0.6	61	22	ABV40905	Peptide #841 enco
269	6	0.6	42	22	AAV68547	Human bone marrow	342	6	0.6	61	22	ABV25034	Protein #7033 enco
270	6	0.6	42	22	AAV14377	Peptide #811 enco	343	6	0.6	61	22	AAV61764	Human brain expres
271	6	0.6	42	22	AAV14394	Peptide #828 enco	344	6	0.6	61	22	AAV74562	Human bone marrow
272	6	0.6	42	22	AAV16354	Peptide #2788 enco	345	6	0.6	61	22	AAV20342	Peptide #6776 enco
273	6	0.6	42	22	AAV26790	Peptide #827 enco	346	6	0.6	61	22	AAV34674	Peptide #8711 enco
274	6	0.6	42	22	AAV26807	Peptide #844 enco	347	6	0.6	61	22	AAV7657	Human colon cancer
275	6	0.6	42	22	AAV28851	Peptide #2888 enco	348	6	0.6	62	22	AAV43437	Propionibacterium
276	6	0.6	42	22	AAV02121	Peptide #803 enco	349	6	0.6	62	22	AAV46812	Propionibacterium
277	6	0.6	42	22	AAV04092	Peptide #2774 enco	350	6	0.6	62	22	ABG09165	Novel human diagno
278	6	0.6	43	22	ABV40314	Peptide #7820 enco	351	6	0.6	63	22	ABV32216	Peptide #4867 enco
279	6	0.6	43	22	ABV24715	Protein #6714 enco	352	6	0.6	63	22	ABV40694	Peptide #8200 enco
280	6	0.6	43	22	AAV61114	Human brain expres	353	6	0.6	63	22	ABV17324	Human nervous syst
281	6	0.6	43	22	AAV73822	Human bone marrow	354	6	0.6	63	22	AAV61556	Human brain expres
282	6	0.6	43	22	AAV20114	Peptide #548 enco	355	6	0.6	63	22	AAV74344	Human bone marrow
283	6	0.6	43	22	AAV34008	Peptide #8045 enco	356	6	0.6	63	22	AAV18426	Peptide #4860 enco
284	6	0.6	44	21	AAV58848	Arabidopsis thalia	357	6	0.6	63	22	AAV34457	Peptide #4894 enco
285	6	0.6	44	22	ABV24649	Novel human diagno	358	6	0.6	63	22	AAV06024	Peptide #4706 enco
286	6	0.6	45	22	AAV85297	Human immune/haema	359	6	0.6	64	22	AAV95255	Human reproductive
287	6	0.6	46	21	AAV38172	Gene 3 human secre	360	6	0.6	65	21	AAV59870	Arabidopsis thalia
288	6	0.6	46	21	AAV38173	Human secreted pro	361	6	0.6	65	22	AAV47323	Propionibacterium
289	6	0.6	48	22	AAV32227	Novel human secret	362	6	0.6	66	21	AAV76173	Human secreted pro
290	6	0.6	49	22	ABV23035	Protein #5034 enco	363	6	0.6	66	22	AAV61207	Propionibacterium
291	6	0.6	49	22	AAE04847	Human SGP014 phosp	364	6	0.6	66	22	AAV66353	Propionibacterium
292	6	0.6	50	19	AAV74753	Human secreted pro	365	6	0.6	66	22	AAV07539	Human polypeptide
293	6	0.6	50	22	AAU42829	Propionibacterium	366	6	0.6	67	15	AAV46071	T-complex like pro
294	6	0.6	51	22	AAV55376	Propionibacterium	367	6	0.6	67	19	AAV20950	Human presentin I
295	6	0.6	52	22	AAV52630	Propionibacterium	368	6	0.6	67	20	AAV12595	Human 5' EST secre
296	6	0.6	52	22	AAV67169	Propionibacterium	369	6	0.6	67	22	ABV27655	Human peptide #306
297	6	0.6	52	22	ABG06727	Novel human diagno	370	6	0.6	67	22	ABV28235	Human peptide #886
298	6	0.6	52	22	ABG08764	Novel human diagno	371	6	0.6	67	22	ABV32825	Peptide #331 enco
299	6	0.6	52	22	ABV15901	Human nervous syst	372	6	0.6	67	22	ABV33410	Peptide #916 enco
300	6	0.6	52	22	AAV20569	Human secreted pro	373	6	0.6	67	22	ABV37859	Peptide #5365 enco
301	6	0.6	53	21	AAV02496	Human secreted pro	374	6	0.6	67	22	ABV18307	Protein #306 enco
302	6	0.6	53	21	AAV65232	Human 5' EST relat	375	6	0.6	67	22	ABV18869	Protein #868 enco
303	6	0.6	53	22	AAV51376	Propionibacterium	376	6	0.6	67	22	ABV22960	Protein #4959 enco

377	6	0.6	67	22	AAW53631	Human brain expres	450	6	0.6	82	20	AAV35247	Chlamydia pneumonia
378	6	0.6	67	22	AAW54195	Human brain expres	451	6	0.6	82	21	AAB52034	Human secreted pro
379	6	0.6	67	22	AAW58490	Human brain expres	452	6	0.6	82	21	AAG07439	Arabidopsis thalia
380	6	0.6	67	22	AAW66011	Human bone marrow	453	6	0.6	82	21	AAW3762	Arabidopsis thalia
381	6	0.6	67	22	AAW66589	Human bone marrow	454	6	0.6	83	21	AAG32727	Zea mays protein f
382	6	0.6	67	22	AAW70990	Human bone marrow	455	6	0.6	84	21	AAW35303	Zea mays protein f
383	6	0.6	67	22	AAW89672	Human Immune/haema	456	6	0.6	85	18	AAW55416	H. pylori ORF hpi
384	6	0.6	67	22	AAW13880	Peptide #314 encod	457	6	0.6	85	22	AAW53552	Proionibacterium
385	6	0.6	67	22	AAW14462	Peptide #896 encod	458	6	0.6	85	22	AAW09985	Human polypeptide
386	6	0.6	67	22	AAW18761	Peptide #5195 enco	459	6	0.6	86	22	AAW46479	B. brevis tyrocid
387	6	0.6	67	22	AAW26287	Peptide #324 encod	460	6	0.6	87	22	AAW52459	Proionibacterium
388	6	0.6	67	22	AAW26875	Peptide #912 encod	461	6	0.6	87	22	AAW52497	Proionibacterium
389	6	0.6	67	22	AAW31266	Peptide #5303 enco	462	6	0.6	88	21	AAW43149	Human ORFX ORF2913
390	6	0.6	67	22	AAW01623	Peptide #305 encod	463	6	0.6	88	22	AAW03843	Human musculoskele
391	6	0.6	67	22	AAW02189	Peptide #871 encod	464	6	0.6	88	22	AAW34895	Human reproductive
392	6	0.6	68	22	AAW83810	Human Immune/haema	465	6	0.6	89	21	AAW26659	Arabidopsis thalia
393	6	0.6	68	22	AAW09990	Human polypeptide/haema	466	6	0.6	89	22	AAW70551	Drosophila melanog
394	6	0.6	69	11	AAW05873	N-terminal sequenc	467	6	0.6	89	22	AAW67108	Human immune/haema
395	6	0.6	69	21	AAW34050	Human secreted pro	468	6	0.6	91	18	AAW27796	Iron (III) dicitra
396	6	0.6	69	21	AAW19806	Arabidopsis thalia	469	6	0.6	91	21	AAW01837	Human secreted pro
397	6	0.6	69	21	AAW33703	Arabidopsis thalia	470	6	0.6	92	20	AAW36565	Fragment of human
398	6	0.6	69	22	AAW33313	Peptide #4964 enco	471	6	0.6	92	21	AAW56482	Human prostate can
399	6	0.6	69	22	AAW22824	Human prostate can	472	6	0.6	93	21	AAW91561	Human secreted pro
400	6	0.6	69	22	AAW96130	Human reproductive	473	6	0.6	93	21	AAW91679	Human secreted pro
401	6	0.6	69	22	AAW58230	Human brain expres	474	6	0.6	94	15	AAW62442	Guanylate cyclase
402	6	0.6	69	22	AAW07169	Human polypeptide	475	6	0.6	94	16	AAW69810	Expression product
403	6	0.6	69	22	AAW30994	Peptide #5031 enco	476	6	0.6	94	22	AAW08920	Human NOVX1 protei
404	6	0.6	70	22	AAW42091	Human ORFX ORF1855	477	6	0.6	94	22	AAW91167	Human immune/haema
405	6	0.6	70	22	AAW60025	Proionibacterium	478	6	0.6	95	11	AAW04198	F gene of simian 1
406	6	0.6	70	22	AAW29249	Peptide #1900 enco	479	6	0.6	95	18	AAW20877	H. pylori cytoplas
407	6	0.6	70	22	AAW34412	Peptide #1918 enco	480	6	0.6	95	22	AAW20451	Arabidopsis thalia
408	6	0.6	70	22	AAW19824	Protein #1823 enco	481	6	0.6	95	22	AAW39861	Proionibacterium
409	6	0.6	70	22	AAW55203	Human brain expres	482	6	0.6	96	21	AAW56643	Partial peptide fr
410	6	0.6	70	22	AAW67597	Human bone marrow	483	6	0.6	97	22	AAW313754	Novel human diagno
411	6	0.6	70	22	AAW15407	Peptide #1841 enco	484	6	0.6	99	10	AAW92035	Sequence encoded 1
412	6	0.6	70	22	AAW27895	Peptide #1932 enco	485	6	0.6	99	22	AAW44904	Proionibacterium
413	6	0.6	70	22	AAW03168	Peptide #1850 enco	486	6	0.6	99	22	AAW47407	Proionibacterium
414	6	0.6	71	22	AAW74469	Human colon cancer	487	6	0.6	99	22	AAW17307	Novel human diagno
415	6	0.6	72	21	AAW87539	Proionibacterium	488	6	0.6	100	19	AAW64533	Human C04 fragment
416	6	0.6	73	22	AAW63992	Novel human diagno	489	6	0.6	100	22	AAW65880	Drosophila melanog
417	6	0.6	73	22	AAW61092	Human polypeptide	490	6	0.6	100	22	AAW05097	Proionibacterium
418	6	0.6	73	22	AAW02183	Human polypeptide	491	6	0.6	100	22	AAW29572	Novel human secret
419	6	0.6	74	22	AAW33702	Arabidopsis thalia	492	6	0.6	101	21	AAW34862	Human secreted pro
420	6	0.6	74	22	AAW87509	Human Immune/haema	493	6	0.6	101	21	AAW21731	Arabidopsis thalia
421	6	0.6	75	21	AAW19653	Arabidopsis thalia	494	6	0.6	101	22	AAW56031	Proionibacterium
422	6	0.6	75	21	AAW00665	Human secreted pro	495	6	0.6	101	22	AAW32073	Novel human secret
423	6	0.6	75	22	AAW52971	Proionibacterium	496	6	0.6	101	22	AAW04124	Human gene 1 encod
424	6	0.6	76	20	AAW73859	Human prostate tum	497	6	0.6	101	22	AAW01825	Human gene 15 enco
425	6	0.6	76	21	AAW14414	Saccharomyces cere	498	6	0.6	101	22	AAW49132	Staphylococcus aur
426	6	0.6	76	22	AAW21167	Human novel foetal	499	6	0.6	102	21	AAW56617	Human prostate can
427	6	0.6	76	22	AAW17906	Novel human respir	500	6	0.6	102	21	AAW40808	Human ORFX ORF572
428	6	0.6	76	22	AAW75570	Human colon cancer	501	6	0.6	102	21	AAW42786	Human ORFX ORF2550
429	6	0.6	77	22	AAW87354	Human immune/haema	502	6	0.6	102	21	AAW44007	Zea mays protein f
430	6	0.6	78	18	AAW55272	H. pylori ORF 05ce	503	6	0.6	103	6	AAW50340	Sequence of sub-un
431	6	0.6	78	22	AAW29525	Peptide #2176 enco	504	6	0.6	103	17	AAW04857	Synthetic cholera
432	6	0.6	78	22	AAW34709	Peptide #2215 enco	505	6	0.6	103	17	AAW06606	Cholera toxin B su
433	6	0.6	78	22	AAW20121	Protein #2120 enco	506	6	0.6	103	17	AAW06607	Cholera toxin B su
434	6	0.6	78	22	AAW55505	Human brain expres	507	6	0.6	103	19	AAW080808	Amino acid sequenc
435	6	0.6	78	22	AAW67890	Human bone marrow	508	6	0.6	103	21	AAW06063	Caenorhabditis ele
436	6	0.6	78	22	AAW15708	Peptide #2142 enco	509	6	0.6	103	21	AAW06130	Caenorhabditis ele
437	6	0.6	78	22	AAW28212	Peptide #2249 enco	510	6	0.6	103	21	AAW18769	Zea mays protein f
438	6	0.6	78	22	AAW03443	Peptide #2125 enco	511	6	0.6	103	22	AAW67871	Proionibacterium
439	6	0.6	79	22	AAW64318	Proionibacterium	512	6	0.6	103	22	AAW62362	V. cholera cholera
440	6	0.6	79	22	AAW00177	Human polypeptide	513	6	0.6	103	22	AAW62364	V. cholera cholera
441	6	0.6	80	21	AAW65667	C. elegans insulin	514	6	0.6	103	22	AAW62365	V. cholera cholera
442	6	0.6	80	22	AAW45074	Arabidopsis thalia	515	6	0.6	103	22	AAW62367	V. cholera cholera
443	6	0.6	81	21	AAW00313	Human secreted pro	516	6	0.6	104	17	AAW04239	Human FKBP12. Hom
444	6	0.6	81	22	AAW49771	Proionibacterium	517	6	0.6	104	19	AAW56026	FK506-binding prot
445	6	0.6	81	22	AAW61059	Human polypeptide	518	6	0.6	104	21	AAW25021	Arabidopsis thalia
446	6	0.6	81	22	AAW01037	Human secreted pro	519	6	0.6	104	21	AAW40961	Zea mays protein f
447	6	0.6	81	22	AAW64202	Subtilopeptidase 9	520	6	0.6	104	21	AAW00653	Human secreted pro
448	6	0.6	81	22	AAW20343	Human secreted pro	521	6	0.6	104	22	AAW15388	Novel human diagno
449	6	0.6	81	22	AAW20343	Human secreted pro	522	6	0.6	104	22	AAW15388	Novel human diagno

523	6	0.6	104	22	ABG16383	Novel human diagno
524	6	0.6	105	22	AAU53993	Propionibacterium
525	6	0.6	105	22	AAU64367	Propionibacterium
526	6	0.6	105	22	ABG10410	Novel human diagno
527	6	0.6	105	22	ABG15259	Novel human diagno
528	6	0.6	105	22	ABG15350	Novel human diagno
529	6	0.6	105	22	AAU16284	Novel human diagno
530	6	0.6	105	22	AAU16284	Novel human diagno
531	6	0.6	106	12	AAU17125	V. cholera cholera
532	6	0.6	106	12	AAU17125	Amino acids 1-106
533	6	0.6	107	12	AAU79307	Human protein SEQ
534	6	0.6	107	12	AAU1581	Macrocyclic FK-506
535	6	0.6	107	16	AAU79857	Bovine FK-506 immu
536	6	0.6	107	16	AAU79857	MDV L1 protein. M
537	6	0.6	107	18	AAU21691	Human FK-binding p
538	6	0.6	107	19	AAU17673	FK506 binding prot
539	6	0.6	107	19	AAU37470	Marek's disease vi
540	6	0.6	107	21	AAU28592	Amino acids 1-107
541	6	0.6	107	22	AAU93339	An Escherichia col
542	6	0.6	107	22	ABU17608	Propionibacterium
543	6	0.6	108	13	AAU20561	Human nervous syst
544	6	0.6	108	13	AAU23008	FKBP. Homo sapien
545	6	0.6	108	14	AAU43692	FK-506 binding pro
546	6	0.6	108	19	AAU61220	Streptococcus pneu
547	6	0.6	108	21	AAU41491	Human ORFX ORF1255
548	6	0.6	108	21	AAU44401	Human FKBP-12 cyto
549	6	0.6	108	22	AAU04561	Human G-protein co
550	6	0.6	108	22	AAU31551	Novel human secret
551	6	0.6	109	22	AAU94914	Human reproductive
552	6	0.6	109	22	AAU603895	Human secreted pro
553	6	0.6	109	22	ABG06434	Novel human diagno
554	6	0.6	109	22	ABG24079	Novel human diagno
555	6	0.6	110	22	AAU93389	Human polypeptide.
556	6	0.6	110	22	AAU01581	Human polypeptide
557	6	0.6	111	21	AAU53250	Human colon cancer
558	6	0.6	112	22	AAU41630	Propionibacterium
559	6	0.6	112	22	AAU001328	Human polypeptide
560	6	0.6	113	22	AAU72085	Glycine max RAD51
561	6	0.6	114	21	AAU26658	Arabidopsis thalia
562	6	0.6	114	22	AAU59848	Propionibacterium
563	6	0.6	114	22	AAU05102	Human polypeptide
564	6	0.6	115	20	AAU42459	Human guanylin pre
565	6	0.6	115	20	AAU42460	Human guanylin pre
566	6	0.6	115	21	AAU12953	Arabidopsis thalia
567	6	0.6	115	21	AAU56949	Arabidopsis thalia
568	6	0.6	115	21	AAU57269	Arabidopsis thalia
569	6	0.6	115	21	AAU02699	Human secreted pro
570	6	0.6	115	22	AAU78557	Human protein SEQ
571	6	0.6	115	22	AAU91675	Human immune/haema
572	6	0.6	115	22	AAU92207	C glutamicum prote
573	6	0.6	116	19	AAU52987	Homo sapiens clone
574	6	0.6	116	20	AAU21592	Human secreted pro
575	6	0.6	116	21	AAU53345	Human colon cancer
576	6	0.6	116	21	AAU26942	Zea mays protein f
577	6	0.6	116	22	AAU82088	Human haematologic
578	6	0.6	117	10	AAU90153	Sequence of hepati
579	6	0.6	117	20	AAU92036	Sequence encoded i
580	6	0.6	117	21	AAU18080	Pinus radiata squa
581	6	0.6	118	11	AAU04163	Cholera toxin B-su
582	6	0.6	118	22	AAU86250	Matize PTE partial
583	6	0.6	118	22	AAU63059	Propionibacterium
584	6	0.6	118	22	AAU10853	Human ovarian and/
585	6	0.6	118	22	AAU94530	Human reproductive
586	6	0.6	118	22	AAU00762	Human polypeptide
587	6	0.6	119	21	AAU34376	Arabidopsis thalia
588	6	0.6	119	22	AAU01960	Human polypeptide
589	6	0.6	119	22	AAU02547	Human polypeptide
590	6	0.6	119	22	AAU06550	Human polypeptide
591	6	0.6	120	19	AAU74835	Human secreted pro
592	6	0.6	120	21	AAU26657	Arabidopsis thalia
593	6	0.6	120	22	AAU23179	Novel human enzyme
594	6	0.6	120	22	AAU01492	Human polypeptide
595	6	0.6	120	22	AAU02771	Human NKp30 recept
596	6	0.6	120	22	AAU21843	Novel human neopla
597	6	0.6	121	22	AAU57757	Propionibacterium
598	6	0.6	121	22	AAU10605	Human macrophage-e
599	6	0.6	121	22	AAU08295	Human PML-1-like pro
600	6	0.6	121	22	AAU31861	Novel human secret
601	6	0.6	121	22	AAU92910	C glutamicum prote
602	6	0.6	121	22	AAU94940	Human protein sequ
603	6	0.6	122	22	AAU36408	Pseudomonas aerugi
604	6	0.6	122	22	AAU80854	Human haematologic
605	6	0.6	122	22	AAU81313	Human haematologic
606	6	0.6	122	22	AAU81872	Human haematologic
607	6	0.6	122	22	AAU00622	Human polypeptide
608	6	0.6	123	18	AAU28331	staphylococcus aur
609	6	0.6	123	22	AAU16477	Human nervous syst
610	6	0.6	123	22	AAU31170	Novel human secret
611	6	0.6	123	22	AAU86760	Human immune/haema
612	6	0.6	124	13	AAU28831	B subunit of CT.
613	6	0.6	124	17	AAU06605	Cholera toxin B su
614	6	0.6	124	21	AAU96872	Plant-optimized V.
615	6	0.6	124	22	AAU65992	Cholera toxin B su
616	6	0.6	124	22	AAU62359	V. cholera strain
617	6	0.6	124	22	AAU62361	V. cholera cholera
618	6	0.6	124	22	AAU62363	V. cholera cholera
619	6	0.6	124	22	AAU62366	V. cholera cholera
620	6	0.6	124	22	AAU62368	V. cholera cholera
621	6	0.6	125	12	AAU33017	Novel human secret
622	6	0.6	126	12	AAU12630	GtB.1/CTB chimeri
623	6	0.6	126	21	AAU01793	Human secreted pro
624	6	0.6	126	22	AAU14324	Novel human diagno
625	6	0.6	126	22	AAU00132	Human polypeptide
626	6	0.6	126	22	AAU73350	Human colon cancer
627	6	0.6	127	20	AAU73985	Human prostate tum
628	6	0.6	127	21	AAU26604	Arabidopsis thalia
629	6	0.6	127	22	AAU03345	Novel human diagno
630	6	0.6	127	22	AAU13520	Novel human diagno
631	6	0.6	128	18	AAU08427	Vector expressing
632	6	0.6	128	20	AAU30055	Amino acid sequenc
633	6	0.6	128	22	AAU86553	Human immune/haema
634	6	0.6	128	22	AAU01856	Human polypeptide
635	6	0.6	129	21	AAU45369	Gene 35 human secr
636	6	0.6	129	21	AAU45370	Human secreted pro
637	6	0.6	129	22	AAU05272	Human polypeptide
638	6	0.6	130	20	AAU29523	Human lung tumour
639	6	0.6	130	20	AAU21597	Human secreted pro
640	6	0.6	130	21	AAU58516	Lung cancer associ
641	6	0.6	130	21	AAU44408	Human lung tumour
642	6	0.6	130	22	AAU13749	Human lung tumour-
643	6	0.6	131	11	AAU04825	lrf-CTB fusion pro
644	6	0.6	131	21	AAU51234	Arabidopsis thalia
645	6	0.6	131	21	AAU56948	Arabidopsis thalia
646	6	0.6	131	21	AAU57268	Arabidopsis thalia
647	6	0.6	131	21	AAU03893	Human secreted pro
648	6	0.6	131	22	AAU70623	Drosophila melanog
649	6	0.6	131	22	AAU25454	Novel human diagno
650	6	0.6	131	22	AAU66714	C glutamicum phosp
651	6	0.6	132	21	AAU32654	Zea mays protein f
652	6	0.6	132	22	AAU19027	Novel human diagno
653	6	0.6	132	22	AAU11519	Human polypeptide
654	6	0.6	133	21	AAU10853	Arabidopsis thalia
655	6	0.6	133	22	AAU84935	Human immune/haema
656	6	0.6	133	22	AAU20155	Segment of human C
657	6	0.6	134	13	AAU04653	Arabidopsis thalia
658	6	0.6	134	21	AAU21826	Arabidopsis thalia
659	6	0.6	134	21	AAU57552	Arabidopsis thalia
660	6	0.6	134	21	AAU57552	Arabidopsis thalia
661	6	0.6	134	22	AAU81669	S. epidermidis ope
662	6	0.6	135	10	AAU90289	Sequence of a fusi
663	6	0.6	135	21	AAU14523	Arabidopsis thalia
664	6	0.6	135	21	AAU20966	Arabidopsis thalia
665	6	0.6	135	21	AAU59173	Arabidopsis thali
666	6	0.6	135	22	AAU02870	CD4 fragment bindi
667	6	0.6	136	21	AAU58752	Human polypeptide
668	6	0.6	136	21	AAU10414	Breast and ovarian
						Arabidopsis thalia

659	6	0.6	136	21	AAG23102	742	6	0.6	150	22	AAU23287	Novel human enzyme
670	6	0.6	136	21	AAG38985	743	6	0.6	150	22	AAAG72988	Olfactory receptor
671	6	0.6	136	22	ABBI0412	744	6	0.6	151	12	AAR13327	HEL epididymis-spe
672	6	0.6	136	22	AAM89699	745	6	0.6	151	18	AAW27292	Human H1075-1 secr
673	6	0.6	136	22	AAM91793	746	6	0.6	151	19	AAW81777	Human HEL protein.
674	6	0.6	136	22	AAO05854	747	6	0.6	151	21	AAAG03731	Human secreted pro
675	6	0.6	136	22	AAU16281	748	6	0.6	151	21	AAAY93597	Protein encoded by
676	6	0.6	137	22	ABB60526	749	6	0.6	151	22	ABBI6399	Human nervous syst
677	6	0.6	137	22	ABG29219	750	6	0.6	151	22	AAW78556	Human protein SEQ
678	6	0.6	137	22	AAM94549	751	6	0.6	152	19	AAW48271	Rat ninturin 1. R
679	6	0.6	137	22	AAO03957	752	6	0.6	152	19	AAW48272	Human ninturin 1.
680	6	0.6	138	18	AAW35844	753	6	0.6	152	22	AAU43525	Protonibacterium
681	6	0.6	138	18	AAW08429	754	6	0.6	152	22	AAW98859	Physcomitrella pat
682	6	0.6	138	21	AAW44858	755	6	0.6	152	22	AAAG92981	C glutamicum prote
683	6	0.6	138	21	AAAY56707	756	6	0.6	155	16	AAW79152	E.coli ribonucleas
684	6	0.6	138	22	AAU57401	757	6	0.6	155	22	ABBI1967	Human secreted pro
685	6	0.6	139	20	AAV35589	758	6	0.6	155	22	AAU30231	Novel human secret
686	6	0.6	139	21	AAAG14522	759	6	0.6	155	22	AAW79540	Human protein SEQ
687	6	0.6	139	21	AAAG20965	760	6	0.6	155	22	AAW79541	Human protein SEQ
688	6	0.6	139	21	AAAG51233	761	6	0.6	155	22	AAO00453	Human polypeptide
689	6	0.6	139	22	ABG09412	762	6	0.6	156	22	AAU36521	Pseudomonas aerugi
690	6	0.6	139	22	ABBA4127	763	6	0.6	156	22	AAU39788	Protonibacterium
691	6	0.6	139	22	AAM61983	764	6	0.6	157	21	AAAG44006	zea mays protein f
692	6	0.6	139	22	AAW74787	765	6	0.6	158	21	AAW18768	zea mays protein f
693	6	0.6	139	22	AAO12894	766	6	0.6	158	22	ABG19100	Novel human diagno
694	6	0.6	139	22	AAM34903	767	6	0.6	159	18	AAW44125	Streptococcus pneu
695	6	0.6	140	21	AAAG14521	768	6	0.6	160	18	AAW11225	S. pneumoniae taga
696	6	0.6	140	21	AAG20964	769	6	0.6	160	18	AAW14722	S. hirsutus GpC.
697	6	0.6	140	22	AAU42645	770	6	0.6	160	20	AAW73637	S. pneumoniae lacC
698	6	0.6	141	17	AAW94409	771	6	0.6	160	21	AAAG26941	Streptococcus pneu
699	6	0.6	141	17	AAW91891	772	6	0.6	160	22	AAW66423	Enterococcus faeca
700	6	0.6	141	19	AAW98794	773	6	0.6	161	20	AAW00081	Enterococcus sp. WF
701	6	0.6	141	19	AAW61433	774	6	0.6	161	20	AAAY00059	Protonibacterium
702	6	0.6	141	20	AAW28500	775	6	0.6	161	21	AAAY97828	Human quanylate Ki
703	6	0.6	141	21	AAAB16393	776	6	0.6	161	22	AAW66010	Human protein sequ
704	6	0.6	141	21	AAAG51232	777	6	0.6	161	22	AAW25783	Human H1075-1 secr
705	6	0.6	141	22	AAU35696	778	6	0.6	162	18	AAW27291	Protonibacterium
706	6	0.6	141	22	AAU35878	779	6	0.6	162	18	AAW36949	Protein encoded by
707	6	0.6	142	18	AAW28124	780	6	0.6	162	22	AAU59012	Protonibacterium
708	6	0.6	142	18	AAW28048	781	6	0.6	163	22	AAU63417	Human quanylate Ki
709	6	0.6	142	22	AAU47085	782	6	0.6	163	22	ABBI2270	Human quanylate Ki
710	6	0.6	143	22	ABG00477	783	6	0.6	163	22	AAW25236	Human quanylate Ki
711	6	0.6	143	22	ABG29995	784	6	0.6	164	21	AAW16379	Human quanylate Ki
712	6	0.6	144	21	ABG35445	785	6	0.6	164	21	AAW88572	Human quanylate Ki
713	6	0.6	144	21	AAAB00123	786	6	0.6	164	21	AAW88376	Human quanylate Ki
714	6	0.6	144	21	AAAG09966	787	6	0.6	164	22	AAU12328	Human quanylate Ki
715	6	0.6	144	21	AAAG50454	788	6	0.6	164	22	AAW20339	Human quanylate Ki
716	6	0.6	144	21	AAAG54790	789	6	0.6	164	22	AAW68597	Human quanylate Ki
717	6	0.6	144	22	AAU54933	790	6	0.6	164	22	AAW80214	Human quanylate Ki
718	6	0.6	144	22	AAW97910	791	6	0.6	165	19	AAW11071	Human quanylate Ki
719	6	0.6	145	18	AAW27812	792	6	0.6	165	21	AAW57227	Human quanylate Ki
720	6	0.6	145	20	AAW37604	793	6	0.6	165	21	AAW37604	Human quanylate Ki
721	6	0.6	145	21	AAW32519	794	6	0.6	165	21	AAW32519	Human quanylate Ki
722	6	0.6	145	21	AAW32519	795	6	0.6	165	21	AAW32519	Human quanylate Ki
723	6	0.6	146	21	AAW56771	796	6	0.6	165	21	AAW56771	Human quanylate Ki
724	6	0.6	146	22	AAW56771	797	6	0.6	166	21	AAW56771	Human quanylate Ki
725	6	0.6	147	22	AAU14415	798	6	0.6	166	22	AAU56780	Human quanylate Ki
726	6	0.6	148	13	AAW20153	799	6	0.6	166	22	AAU58218	Human quanylate Ki
727	6	0.6	148	22	AAU29520	800	6	0.6	166	22	AAU29562	Human quanylate Ki
728	6	0.6	149	21	AAW57795	801	6	0.6	166	22	AAW57795	Human quanylate Ki
729	6	0.6	149	22	ABG18726	802	6	0.6	167	21	AAW16491	Human quanylate Ki
730	6	0.6	149	22	ABG25600	803	6	0.6	167	21	AAW37995	Human quanylate Ki
731	6	0.6	149	22	AAW90245	804	6	0.6	167	21	AAW38326	Human quanylate Ki
732	6	0.6	149	22	AAW90245	805	6	0.6	167	21	AAW90245	Human quanylate Ki
733	6	0.6	150	18	AAW23817	806	6	0.6	167	22	AAW23817	Human quanylate Ki
734	6	0.6	150	19	AAW69326	807	6	0.6	168	21	AAW69326	Human quanylate Ki
735	6	0.6	150	19	AAW31115	808	6	0.6	168	21	AAW31115	Human quanylate Ki
736	6	0.6	150	19	AAW33333	809	6	0.6	168	22	AAW33333	Human quanylate Ki
737	6	0.6	150	19	AAW39137	810	6	0.6	168	22	AAW39137	Human quanylate Ki
738	6	0.6	150	21	AAW19247	811	6	0.6	169	21	AAW19247	Human quanylate Ki
739	6	0.6	150	21	AAW67556	812	6	0.6	169	21	AAW67556	Human quanylate Ki
740	6	0.6	150	21	AAW78889	813	6	0.6	169	21	AAW78889	Human quanylate Ki
741	6	0.6	150	22	ABG23089	814	6	0.6	169	21	AAW23089	Human quanylate Ki

815	6	0.6	169	22	AAE03810	Protein encoded by
816	6	0.6	169	22	AAE03811	Protein encoded by
817	6	0.6	170	21	AAAB40683	Human ORFX ORF447
818	6	0.6	170	21	AAAG57794	Arabidopsis thalia
819	6	0.6	171	20	AAAV37016	Protein Involved I
820	6	0.6	171	21	AAAG15743	Arabidopsis thalia
821	6	0.6	171	21	AAAG37250	Arabidopsis thalia
822	6	0.6	171	22	AAAM82139	Human haematologic
823	6	0.6	172	22	AAU64154	Protonibacterium
824	6	0.6	172	22	AAAG23495	Novel human diagno
825	6	0.6	172	22	AAU31796	Novel human secreto
826	6	0.6	172	22	AAAB9670	Schizosaccharomyce
827	6	0.6	173	10	AAAP90740	Sequence of a fusi
828	6	0.6	173	21	AAAY59174	CD4 fragment bindi
829	6	0.6	173	22	AAAG72893	Olfactory receptor
830	6	0.6	173	22	AAAB94167	Human protein sequ
831	6	0.6	174	21	AAAG44654	Zea mays protein f
832	6	0.6	174	22	AAAG05983	Novel human diagno
833	6	0.6	174	22	AAAG29097	Novel human diagno
834	6	0.6	175	22	AAAB92657	Human protein sequ
835	6	0.6	176	11	AAAR07544	Segment comprising
836	6	0.6	176	13	AAAR27283	Sequence of chimae
837	6	0.6	176	22	AAAB67332	Protonibacterium
838	6	0.6	177	20	AAAY06402	Human B-cell myeli
839	6	0.6	177	21	AAAG21868	Arabidopsis thalia
840	6	0.6	179	21	AAAB58191	Lung cancer associ
841	6	0.6	179	22	AAU58316	Protonibacterium
842	6	0.6	179	22	AAU23732	Novel human diagno
843	6	0.6	181	18	AAAM28332	Staphylococcus aur
844	6	0.6	181	21	AAAG04887	Arabidopsis thalia
845	6	0.6	181	21	AAAG24433	Arabidopsis thalia
846	6	0.6	181	21	AAAG37998	Arabidopsis thalia
847	6	0.6	181	22	AAU17436	Novel signal trans
848	6	0.6	181	22	AAU21805	Novel human neopla
849	6	0.6	184	21	AAAB27975	Human secreted pro
850	6	0.6	184	21	AAAB27976	Human secreted pro
851	6	0.6	185	20	AAAY14206	HIV-1 gp120 peptid
852	6	0.6	185	20	AAAY14210	HIV gp120 protein
853	6	0.6	185	21	AAAG33624	Arabidopsis thalia
854	6	0.6	185	21	AAAY58587	Sorangium cellulos
855	6	0.6	185	22	AAAB52947	Escherichia coli p
856	6	0.6	185	22	AAAB52981	Escherichia coli p
857	6	0.6	186	10	AAAP90616	DNA mismatch repai
858	6	0.6	186	17	AAAP92522	Fish growth hormo
859	6	0.6	187	9	AAAP81244	Sequence of yellow
860	6	0.6	187	19	AAAM71372	Synthetic fish gro
861	6	0.6	187	21	AAAB41409	Human ORFX ORF173
862	6	0.6	187	22	AAU41292	Protonibacterium
863	6	0.6	187	22	AAAG17972	Novel human diagno
864	6	0.6	188	11	AAAR07353	Fish growth hormo
865	6	0.6	188	13	AAAR24137	Fish growth hormo
866	6	0.6	188	21	AAAG07502	Arabidopsis thalia
867	6	0.6	188	21	AAAG33623	Arabidopsis thalia
868	6	0.6	188	21	AAAG52559	Arabidopsis thalia
869	6	0.6	188	22	AAAG05834	Novel human diagno
870	6	0.6	188	22	AAU19640	Human novel extrac
871	6	0.6	189	20	AAAY00080	Enterococcus faeca
872	6	0.6	189	20	AAAY00058	Enterococcus faeca
873	6	0.6	190	20	AAAY06401	Human B-cell myeli
874	6	0.6	190	21	AAAG04878	Arabidopsis thalia
875	6	0.6	190	21	AAAG29671	Arabidopsis thalia
876	6	0.6	190	22	AAE02769	Human NKp30 recept
877	6	0.6	191	12	AAAR11726	Pre-pro AMF/CD4 fu
878	6	0.6	191	21	AAAG16895	Arabidopsis thalia
879	6	0.6	191	21	AAAG33622	Arabidopsis thalia
880	6	0.6	191	22	AAAB57825	Drosophila melanog
881	6	0.6	191	22	AAAG16555	Protonibacterium
882	6	0.6	191	22	AAE03806	Rat interferon-lik
883	6	0.6	192	18	AAAM20786	H. pylori cytoplas
884	6	0.6	192	22	AAAG09662	Human polypeptide
885	6	0.6	193	21	AAAB40498	Human ORFX ORF262
886	6	0.6	193	22	AAU36078	Klebsiella pneumon
887	6	0.6	193	22	AAAG14727	Novel human diagno
888	6	0.6	194	22	AAAG03338	Novel human diagno
889	6	0.6	195	22	AAAG72975	Olfactory receptor
890	6	0.6	195	22	AAAG10413	Arabidopsis thalia
891	6	0.6	196	21	AAAG23101	Arabidopsis thalia
892	6	0.6	196	21	AAAG53517	Arabidopsis thalia
893	6	0.6	197	22	AAAG67254	Amino acid sequenc
894	6	0.6	197	22	AAAB94495	Human protein sequ
895	6	0.6	198	19	AAAM43082	Mycothacterial hepa
896	6	0.6	198	19	AAAM44936	Mycothacterial hepa
897	6	0.6	198	19	AAAG14984	Arabidopsis thalia
898	6	0.6	198	22	AAAB67976	Drosophila melanog
899	6	0.6	198	22	AAAB63319	Human breast cance
900	6	0.6	199	20	AAAY34608	Chlamydia pneumoni
901	6	0.6	199	22	AAU37135	Staphylococcus aur
902	6	0.6	199	22	AAU64689	Protonibacterium
903	6	0.6	200	10	AAAP90732	Fish growth hormo
904	6	0.6	200	16	AAAR82900	Mouse B7-1 (IgV-1)
905	6	0.6	200	21	AAAG10833	Arabidopsis thalia
906	6	0.6	200	21	AAAG27384	Arabidopsis thalia
907	6	0.6	200	21	AAAG50909	Arabidopsis thalia
908	6	0.6	201	20	AAAY06403	Human B-cell myeli
909	6	0.6	201	21	AAAG10412	Arabidopsis thalia
910	6	0.6	201	21	AAAG15683	Human protein sequ
911	6	0.6	201	22	AAAM78446	Fish growth hormo
912	6	0.6	203	10	AAAP90617	Fish growth hormo
913	6	0.6	203	11	AAAR07248	CD4 domains D1-D2.
914	6	0.6	203	16	AAAR78674	CD4 domains D1-D2.
915	6	0.6	203	17	AAAR89451	CD4 D1-D2 domains.
916	6	0.6	203	21	AAAG44547	Arabidopsis thalia
917	6	0.6	203	21	AAAG07637	Novel human diagno
918	6	0.6	203	22	AAU32562	Novel human secreto
919	6	0.6	203	22	AAU21680	Novel human neopla
920	6	0.6	204	10	AAAP95638	Fish growth hormo
921	6	0.6	204	12	AAAR10912	Fish growth hormo
922	6	0.6	204	18	AAAM25120	Yellowtail tuna gr
923	6	0.6	204	18	AAAM27340	Yellow tail/fin tu
924	6	0.6	204	20	AAAM88327	Acetyl transferase
925	6	0.6	204	22	AAAB71729	Drosophila melanog
926	6	0.6	204	22	AAU53997	Protonibacterium
927	6	0.6	204	22	AAAB11429	Human transmembran
928	6	0.6	204	22	AAAM80112	S. epidermidis ope
929	6	0.6	204	22	AAAG81659	S. epidermidis ope
930	6	0.6	204	22	AAAG82062	MS-15 lissue-pref
931	6	0.6	205	19	AAAM46903	Human polypeptide
932	6	0.6	205	22	AAAM99922	Malassezia fungus
933	6	0.6	206	18	AAAM29770	Drosophila melanog
934	6	0.6	206	22	AAAB58142	Arabidopsis thalia
935	6	0.6	207	21	AAAG42951	Arabidopsis thalia
936	6	0.6	208	21	AAAG10832	Arabidopsis thalia
937	6	0.6	208	21	AAAG27383	Arabidopsis thalia
938	6	0.6	208	21	AAAG50908	Arabidopsis thalia
939	6	0.6	208	22	AAAG21397	Novel human diagno
940	6	0.6	208	22	AAAU00853	S. aureus conserved
941	6	0.6	209	14	AAAR34708	Prod. of the AcetI
942	6	0.6	209	19	AAAM68581	Human COP9 protein
943	6	0.6	209	21	AAAG04117	Arabidopsis thalia
944	6	0.6	209	21	AAAG06892	Arabidopsis thalia
945	6	0.6	209	21	AAAG11858	Arabidopsis thalia
946	6	0.6	209	21	AAAG54004	Arabidopsis thalia
947	6	0.6	210	21	AAAG15534	Arabidopsis thalia
948	6	0.6	210	21	AAAG17748	Arabidopsis thalia
949	6	0.6	210	21	AAAG61007	Arabidopsis thalia
950	6	0.6	210	22	AAU67465	Protonibacterium
951	6	0.6	210	22	AAU29449	Human G protein-co
952	6	0.6	211	15	AAAR62661	Hepatitis type-C v
953	6	0.6	211	15	AAAY11102	H. pylori ORF 14ce
954	6	0.6	211	21	AAAG10831	Arabidopsis thalia
955	6	0.6	211	21	AAAG15742	Arabidopsis thalia
956	6	0.6	211	21	AAAG27382	Arabidopsis thalia
957	6	0.6	211	21	AAAG37249	Arabidopsis thalia
958	6	0.6	211	21	AAAG50507	Arabidopsis thalia
959	6	0.6	212	20	AAAY30078	Protein encoded by
960	6	0.6	212	20	AAAY28994	Amino acid sequenc

961	6	0.6	2.12	21	AAAB27945	Human secreted protein
962	6	0.6	2.12	22	AAU37356	Staphylococcus aureus
963	6	0.6	2.12	22	AAU37504	Staphylococcus aureus
964	6	0.6	2.12	22	AAAB95582	Human protein sequ
965	6	0.6	2.12	22	AAAB36585	Human FLEXIT-7 protein
966	6	0.6	2.13	19	AAV11072	H. pylori ORF hp2p
967	6	0.6	2.13	21	AAAG47012	Arabidopsis thaliana
968	6	0.6	2.14	16	AAAB82801	Mouse B7-1 (IgV-11)
969	6	0.6	2.14	21	AAAG20979	Arabidopsis thaliana
970	6	0.6	2.16	20	AAAG35337	A. thaliana CBP3 p
971	6	0.6	2.16	21	AAAG44176	Arabidopsis thaliana
972	6	0.6	2.16	21	AAAB03745	Dehydration respon
973	6	0.6	2.16	21	AAAB03753	Dehydration respon
974	6	0.6	2.16	21	AAAB82475	A. thaliana transcr
975	6	0.6	2.16	22	AAAB25866	Human protein sequ
976	6	0.6	2.16	22	AAE02569	A. thaliana transcr
977	6	0.6	2.17	12	AAAB15150	PCD4-gelsolin sequ
978	6	0.6	2.18	22	AAAB93195	Human polypeptide,
979	6	0.6	2.18	22	AAAB81734	S. epidermidis opo
980	6	0.6	2.19	19	AAAB08693	S. pneumoniae DNA
981	6	0.6	2.20	19	AAAB79323	Staphylococcus aur
982	6	0.6	2.20	20	AAAB35523	Chlamydia pneumoni
983	6	0.6	2.20	21	AAAB38499	Fragment of human
984	6	0.6	2.20	21	AAAG16490	Arabidopsis thalia
985	6	0.6	2.20	21	AAAG37994	Arabidopsis thalia
986	6	0.6	2.21	20	AAAB36863	Protein which is s
987	6	0.6	2.21	21	AAAG21867	Arabidopsis thalia
988	6	0.6	2.22	21	AAAY51400	S. pombe Weel cata
989	6	0.6	2.22	22	AAAB62518	S. pombe Weel prot
990	6	0.6	2.23	13	AAAB29923	SAP. Homo sapiens
991	6	0.6	2.23	21	AAAG06620	Arabidopsis thalia
992	6	0.6	2.23	21	AAAG14149	Arabidopsis thalia
993	6	0.6	2.23	21	AAAG38954	Arabidopsis thalia
994	6	0.6	2.23	22	AAAG05804	Novel human diagn
995	6	0.6	2.24	13	AAAB25675	Monoclonal antibod
996	6	0.6	2.24	18	AAAM14788	FKBP-1CR:SH2 fusio
997	6	0.6	2.24	20	AAAB96823	A fusion protein o
998	6	0.6	2.24	21	AAAG54892	Human protein clon
999	6	0.6	2.24	22	AAAB15369	Novel human diagn
1000	6	0.6	2.24	22	AAE02549	A. thaliana transcr

ALIGNMENTS

RESULT	1
AAAM79303	
ID	AAAM79303 standard; Protein; 502 AA.
XX	
AC	AAAM79303;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human protein SEQ ID NO 1965.
XX	
KM	Human; cytokine; cell proliferation; cell differentiation; gene therapy
KM	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM	tissue; growth factor; immunomodulatory; cancer; leukaemia;
KM	nervous system disorder; arthritis; inflammation.
XX	
OS	Homo sapiens.
XX	
PN	W0200157190-A2.
XX	
PD	09-AUG-2001.
XX	
PF	05-FEB-2001; 2001MO-US04098.
XX	
PR	03-FEB-2000; 2000US-0496914.
PR	27-APR-2000; 2000US-0560875.
PR	20-JUN-2000; 2000US-0598075.
PR	19-JUL-2000; 2000US-0620325.
PR	01-SEP-2000; 2000US-0654936.

15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Meng D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK52436.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 4376-4377; 6221bp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoietic regulatory
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAW80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX
XX Sequence 502 AA;

	Query Match	Similarity	34.2%	Score 367	DB 22	Length 502	
	Match Local	Similarity	100.0%	Pred. No. 0			
	Matches 367	Conservative	0	Mismatches	0	Indels	0
						Gaps	0
Qy	708	MCRILNDEAAMWELARACLNHMEVEFFAIRVRRIGNVGLVMSLEQIKGIEDYNLLAGHILA	767				
Db	1	mcrlindaeamwelaacclhnevefaiirvrrignvgvlmsleqikgiedynllaghta	60				
Qy	768	MFTNDYNLAQDLYASSCPDIAALEKRRDLQHWDSALQLAHLAPDIPRISKERYALQLEF	827				
Db	61	mtfndynlaqdllyasscpdlaalemrtrdlqhwdsalqlahlapdipriskyeyalqlef	120				
Qy	828	AGDYNNALAHYKGTGTGONKKEHDEACLAGVAQMSIRMGDIRGVNQALNHRPRLVKRDCG	887				
Db	121	agdynnalahyekgtitgdhkehdaclagvaqmslrmgdlrtgvnqalnhpsrvlkrdcg	180				
Qy	888	ALLENMKOFSESAOLYEEKGLVYDKAASVYIRSKNNAKVGDLLPRHVSPPRIHQYAKAKEA	947				
Db	181	allemkqfseaaqlyekglvydkaasvyrsknnakvgdllprhvspprlhgyakakea	240				
Qy	948	DGRYKEAVVAYENAKQOMSVIRIYLDHLNPNPEKAVNIARETOSLDGAKVVAFFLQLDGY	1007				
Db	241	dgrypeavvayenakqgsviriyldhlnpnekavniretqslgdakmvaffliqldgy	300				
Qy	1008	GSATGFLVMSKNCNNATFLTAOQHMKMEYIADITIGSEDTTNEDYOSALYFEEGKRYLQWG	1067				
Db	301	gsatgflvmskncneafltaqghmkmeiyaditigsedttneedygsialyfeegkryldqg	360				
Qy	1068	KFFELCG 1074					
Db	361	kffllcg 367					

RESULT 2

AA80287

ID AA80287 standard; Protein: 515 AA.

XX

AC	AAAM80287;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human protein SEQ ID NO 3933.
XX	
KM	Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KM	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM	tissue growth factor; immunomodulatory; cancer; leukaemia;
XX	nervous system disorder; arthritis; inflammation.
OS	Homo sapiens.
XX	
PN	WO200157190-A2.
PD	
XX	09-AUG-2001.
PF	
XX	05-FEB-2001; 2001WO-US04098.
PR	
XX	03-FEB-2000; 2000US-0496914.
PR	27-APR-2000; 2000US-0560875.
PR	20-JUN-2000; 2000US-0598075.
PR	19-JUL-2000; 2000US-0620325.
PR	01-SEP-2000; 2000US-0654935.
PR	15-SEP-2000; 2000US-0663561.
PR	20-OCT-2000; 2000US-0693325.
PR	30-NOV-2000; 2000US-0728422.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Dimaec RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI	Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;
DR	WPI: 2001-476283/51.
XX	N-PSTDB: AAK53420.
PT	Nucleic acids encoding polypeptides wltch cytokine-like activities,
PT	useful in diagnosis and gene therapy -
XX	
PS	Claim 20: Page 471; 6221pp: English.
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC	encoded polypeptides (AAM78323-AAH80302) that exhibit activity elating to
CC	cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g. stem cell growth factor activity, haematopoesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation.
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC	(AAH80020) are omitted as the relevant pages from the sequence listing
CC	were missing at the time of publication.
XX	
SQ	Sequence 515 AA:
Query Match	25.8%; Score 277; DB 22; Length 515;
Best Local Similarity	100.0%; Pred. NO. 1.5e+280;
Matches 277; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	7 HHDSALQLKLHAPDQIPISKEVATOLEFAGDYVALAHYEGTIGDKHEDEACTAGV 857 7 hmdsaqlakhlapdqipitskeyaqifafagdyvalahyegltvgdnkendaclagv 66
Oy	AOMSIKMGDIRGIVGNALNHPRSVLKRDCGALLENNKOPSEAAOLYEKGLYDKAASYI 917 67 agmslrmgtirrtgvgnalhpsrvtlkrdcgallemmkxfseaaqlyekglydkaasyi 126
Oy	RSKNAKKYGDDLPHVSSPFIHQYAKAKADGRKYEAUVAYENAKOWMSVIRIYLDHLNN 977

Db 127 rsknkavqdl1phvssppknhlgayakaeagrykveawenakqgsvirrl10hlm 186
Qy 978 PEKAVNIIVETQSLDCAKMAVAREFLDGDVSAIOFLWMSKCNNEAFTLQAQHNKMEIYA 1037
Db 187 pekavniivretqslgdakmarffllgldgysaigqlfwmkskneatflaqghnkmeiya 246
Qy 1038 DIIGSEDTTNEGYOSTIALYFEGEKRRLLQAGKFFLLCG 1074
Db 247 diigsedttnegyostialyfegekrrllqagkffllcg 283

RESULT 3
AAU14751
ID AAU14751 standard; Protein; 138 AA.
AC AAU14751;
XX
XX
XX 24-OCT-2001 (first entry)
DT
XX
XX Novel bone marrow polypeptide #150.
DE
XX
XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KM haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KM wound healing; nutritional supplement; immune disorder;
KM severe combined immunodeficiency; SCID.
XX
XX Homo sapiens.
XX
XX WO200157187-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 05-FEB-2001; 2001WO-US03782.
XX
XX
XX 03-FEB-2000; 2000US-0496914.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 30-NOV-2000; 2000US-0250683.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
PI
XX
XX WPI: 2001-488875/53.
XX N-PSDB; AAS23056.
DR
XX
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT and gene therapy.
XX
XX
XX Claim 10; Page 135; 392pp; English.

AAU14602-AAU14794 represent novel bone marrow polypeptides of the
invention. The proteins and corresponding coding sequences may be used
in the prevention, diagnosis and treatment of diseases associated with
inappropriate bone marrow polypeptide expression. For example, to treat
disorders associated with decreased expression by rectifying mutations
or deletions in a patient's genome that affect the activity of the
polypeptides by expressing inactive proteins or to supplement the
patient's own production of the polypeptide. Additionally, the nucleic
acids may be used to produce the polypeptides, by inserting the nucleic
acids into a host cell and culturing the cell to express the protein.
The nucleic acid and its complementary sequences may also be used as DNA
probes in diagnostic assays to detect and quantitate the presence of
similar nucleic acid sequences in samples, and therefore which patients
may be in need of restorative therapy. The proteins may also be used as
antigens in the production of antibodies against bone marrow proteins
and in assays to identify modulators of their expression and activity.
The anti-bone marrow protein antibodies and antagonists may also be used
to down regulate expression and activity. The antibodies may also be used
as diagnostic agents for detecting the presence of the protein in samples
(e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins

CC may be used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ Sequence 138 AA;

Query Match 7.4%; Score 80; DB 22; Length 138;
Best Local Similarity 100.0%; Pred. No. 6.3e-75;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 NNOFLMKDDRTSAESHSIVYLKKTLPFLNLEPPNADLEPQDGFNIVCNWGD 282
Db 59 nmqflmkddrtsaeshsivylgkklflnleppnadlefgdfignivywygd 118
|||||

OY 283 GRMIGSGHFVISTHNG 302
Db 119 grmigsgcghfvisthng 138
|||||

RESULT 4
AAU14657
ID AAU14657 standard; Protein; 244 AA.
XX
AC AAU14657;
XX

DT 24-OCT-2001 (first entry)
XX

DE Novel bone marrow polypeptide #56.
XX

KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.
XX

OS Homo sapiens.
XX

PN WO200157187-A2.
XX

PD 09-AUG-2001.
XX

PF 05-FEB-2001; 2001WO-US03782.
XX

PR 03-FEB-2000; 2000US-0496914.
XX

PR 20-JUN-2000; 2000US-0598075.
XX

PR 19-JUL-2000; 2000US-0620325.
XX

PR 30-NOV-2000; 2000US-0250683.
XX

PA (HYSE-) HYSEQ INC.
XX

PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
XX Ren F, Drmanac RT;
XX WPI: 2001-488875/53.
XX DR N-PSDB; AAS22962.
XX

PT Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
XX and gene therapy -
XX

PS Claim 10; Page 239-240; 392pp; English.
XX

AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequences may also be used as DNA

CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins
CC and in assays to identify modulators of their expression and activity.
CC The anti-bone marrow protein antibodies and antagonists may also be used
CC to down regulate expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins
CC may be used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ Sequence 244 AA;

Query Match 5.8%; Score 62; DB 22; Length 244;
Best Local Similarity 100.0%; Pred. No. 7.3e-36;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 MDMDKGDVLAIVAKSSCIYLMDFNFKTSOLDNGMRDMSFLMSKVSFLAVGTGK 159
Db 1 mdwdkdgdvliavaekssciylwdanfnktsqldngmrdmsflmskvsflavgtvgk 60
|||||

OY 160 NL 161
Db 61 nl 62
||

RESULT 5
AAU33013
ID AAU33013 standard; Protein; 72 AA.
XX
AC AAU33013;
XX

DT 18-DEC-2001 (first entry)
XX

DE Novel human secreted protein #3504.
XX

KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX

OS Homo sapiens.
XX

PN WO200179449-A2.
XX

PD 25-OCT-2001.
XX

PF 16-APR-2001; 2001WO-US08656.
XX

PR 18-APR-2000; 2000US-0552929.
XX

PR 26-JAN-2001; 2001US-0770160.
XX

PA (HYSE-) HYSEQ INC.
XX

PI Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-611725/70.
XX

PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy -
XX

PS Claim 20; Page 698; 765pp; English.
XX

The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or

CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.

XX Sequence 72 AA:

Query Match 3.0%; Score 32; DB 22; Length 72;
Best Local Similarity 100.0%; Pred. No. 5.8e-25;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 295 VVISTHTGELGGEIFQARNHKNLTSIAVSQT 326
|||||
DB 29 vvishtgtelgqelfgarhknltslavsq 60

RESULT 6

AAU374711
ID AAG74711 standard; Protein: 237 AA.

XX AAG74711;

AC 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:5475.

KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma; chromosome 1.

XX Homo sapiens.

PN WO200122920-A2.

PD 05-APR-2001.

PF 28-SEP-2000; 2000WO-US26524.

PR 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

PT WPI: 2001-235357/24.

DR N-PSDB: AAH34116.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11: Page 7076-7077; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 237 AA:

Query Match 2.3%; Score 25; DB 22; Length 237;
Best Local Similarity 100.0%; Pred. No. 3.7e-17;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1050 YOSIALYFEGEKRYIQAGKFFLLCG 1074
|||||
DB 1 ysiailyfegekryiqagkffllcg 25

RESULT 7

AAU33123
ID AAU33123 standard; Protein: 109 AA.

XX AAU33123;

DE 18-DEC-2001 (first entry)

DE Novel human secreted protein #3614.

KW Human: vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.

PN WO200179449-A2.

PD 25-OCT-2001.

PF 16-APR-2001; 2001WO-US08656.

PR 18-APR-2000; 2000US-0552929.

PR 26-JAN-2001; 2001US-0770160.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

PT WPI: 2001-611725/70.

DR Nucleic acids encoding a range of human polypeptides, useful in genetic

PT vaccination, testing and therapy -

XX Claim 20: Page 710; 765pp; English.

CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.

XX Sequence 109 AA:

Query Match 1.8%; Score 19; DB 22; Length 109;
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1032 KMEIYADITIGSEDTNEDY 1050
|||||
DB 72 KMEIYADITIGSEDTNEDY 90

RESULT 8

AAV41351
ID AAV41351 standard; Protein: 55 AA.

AAV41351;

02-DEC-1999 (first entry)

Human secreted protein encoded by gene 44 clone HMXEY51.

Human; secreted protein; fusion protein; gene therapy; protein therapy;
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
developmental abnormality; foetal deficiency; blood; allergy; renal;
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
inflammation; ischaemic shock; Alzheimer's disease; testostosis; AIDS;
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

Homo sapiens.

WO9947540-A1.

23-SEP-1999.

18-MAR-1999; 99WO-US05804.

19-MAR-1998; 98US-0078563.

19-MAR-1998; 98US-0078566.

19-MAR-1998; 98US-0078573.

19-MAR-1998; 98US-0078574.

19-MAR-1998; 98US-0078576.

19-MAR-1998; 98US-0078577.

19-MAR-1998; 98US-0078578.

19-MAR-1998; 98US-0078579.

19-MAR-1998; 98US-0078581.

01-APR-1998; 98US-0080312.

01-APR-1998; 98US-0080313.

01-APR-1998; 98US-0080314.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;

Wel Y, Endress CA, Duan RD, Kyaw H, Ebner R, Lafleur DW;

Olsen HS, Shi Y, Moore PA;

WPI; 1999-562050/47.

N-PSDB; AA224854.

New isolated human genes, useful for diagnosis and treatment of e.g.

cancers, neurological disorders, immune diseases, inflammation or blood

disorders -

Claim 11; Page 385; 484pp; English.

This sequence represents a secreted human protein encoded by the gene

clone detailed in the descriptor line. The gene can be used to generate

fusion proteins by linking to the gene to a human immunoglobulin Fc

portion (e.g. AA224802) for increasing the stability of the fused

protein as compared to the human protein only.

The invention relates to 95 novel genes and their fragments (nucleic

acid sequences: AA224811-224907; amino acid sequences AAV41308-141404)

CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 95
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AA224811 for described uses).

SQ Sequence 55 AA;

Query Match 0.7%; Score 7; DB 20; Length 55;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 873 GALKHPS 879
|||||
DB 37 GALKHPS 43

RESULT 9

AA38501
ID AAR38501 standard; protein: 57 AA.

AA38501;

28-OCT-1993 (first entry)

M. bovis pili protein submolecular unit.

Antibodies; whole pili binding; basis; vaccine; bacterial infection;

ruminant footrot infection; sheep; type IV pillated bacteria.

Moraxella bovis.

WO9311791-A.

24-JUN-1993.

17-DEC-1992; 92WO-US11085.

18-DEC-1991; 91US-0809762.

(UOR-) UNIV OREGON HEALTH SCI.

Smith AW;

WPI; 1993-213824/26.

Antigenic preparation - stimulates production of antibodies binding

to pili protein of type IV pillated bacteria, useful in

vaccine compsn.

Claim 12; Page 26; 44pp; English.

The sequence is that of a submolecular unit of Moraxella bovis

pilin protein which corresponds to at least one epitope common to

structural pilin proteins of type IV pillated bacteria. It is

capable of eliciting antibodies (Abs) which bind to whole pili of

type IV bacteria. The ability of the submolecular unit to produce

Abs which bind to whole pili provides the basis for vaccines against

type IV bacterial infections, e.g. footrot infection in ruminants.

SQ Sequence 57 AA;

Query Match 0.7%; Score 7; DB 14; Length 57;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 481 AALFEK 487
|||||

DB 49 aalfeq 55

RESULT 10
 AAB48729
 ID AAB48729 standard; Protein; 58 AA.
 XX
 AC AAB48729;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Mouse liver growth hormone-induced clone 5 ORF #6.
 XX
 KW Mouse: growth hormone; GH regulatable gene; liver pathology; hypertrophy;
 KW hepatocellular lesion; hyperplasia; altered expression level; clone 5;
 KW diagnostic marker; gigantism; acromegaly; diabetes; hepatotropic;
 KW transgenic animal; drug screening; drug discovery; murine;
 KW open reading frame; ORF.
 XX
 OS Mus sp.
 XX
 PN W0200066787-A2.
 XX
 PD 09-NOV-2000.
 XX
 PF 05-MAY-2000; 2000MO-US12366.
 XX
 PR 05-MAY-1999; 99US-0132663.
 XX
 PA (UYOH-) UNIV OHIO.
 XX
 PI Kopschick JJ, Tlong J;
 XX
 DR WPI: 2001-007239/01.
 DR N-PSDB: AAC87257.
 XX
 PT Diagnosing abnormal levels of growth hormone activity in liver
 PT comprising assaying growth transcriptional activity and protein
 PT expression level of hormone-regulatable liver genes, as diagnostic
 PT markers of liver pathology -
 XX
 PS Example 2; Page 53; 65pp; English.
 XX
 CC The invention relates to a method of diagnosing abnormal levels of
 CC growth hormone (GH) activity in the liver, or predicting a change in the
 CC condition of the liver in response to abnormal GH activity. The method
 CC involves correlating the level of expression of certain specific genes
 CC with the level of GH activity in the liver, or with an expected change
 CC in the condition of the liver as the result of GH activity. Excessive GH
 CC activity in the liver is thought to be deleterious to health, causing an
 CC increase in liver size as a consequence of both hyperplasia and
 CC hepatocyte hypertrophy, and hepatocellular lesions which progress with
 CC age. Studies in transgenic mice which express high levels of bovine
 CC growth hormone identified a number of genes whose expression in the
 CC liver is altered by high GH levels. The genes which are upregulated are
 CC those encoding alpha-fetoprotein, corticosteroid binding globulin,
 CC fetuin, ras8-interacting protein, paroxonase-3, cytochrome P450I1A,
 CC 5-2 hydroxyacid oxidase, interferon alpha/beta receptor, growth hormone
 CC receptor, proteasome 2-subunit, and coagulation factor V. Two novel
 CC genes, clone 5 (AAC87257) and clone 45 (AAC87258) are also upregulated
 CC in response to abnormally high GH levels. Conversely, expression of the
 CC gene encoding 3-beta-hydroxysteroid dehydrogenase/delta-5-delta-4
 CC isomerase appears to be downregulated. The invention also relates to
 CC transgenic nonhuman mammals comprising a GH-induced transgene which
 CC exhibit or have a propensity to develop a liver pathology; an assay for
 CC drugs which inhibit the development of, or which treat a liver pathology,
 CC comprising administering the drug to the transgenic animal; and
 CC preventing or treating a liver pathology in a patient comprising
 CC administering a drug which inhibits the expression of a GH-regulatable
 CC gene. The method of the invention is used for diagnosing abnormal levels of GH
 CC activity in the liver or predicting a change in the condition of the
 CC liver in response to abnormal levels of GH activity. The GH-regulatable
 CC liver genes and proteins are useful as diagnostic markers of liver
 CC pathology. Assays for the expression of these genes is useful for the

CC diagnosis of liver pathologies associated with gigantism or acromegaly or
 CC with diabetes, as other causative agents may act directly or indirectly
 CC upon the same genes. Sequence AAB48724-BA8739 represent the open reading
 CC frames (ORFs) of the novel mouse liver cDNA clone 5.
 XX
 XX
 SQ Sequence 58 AA;
 XX
 Query Match 0.7%; Score 7; DB 22; Length 58;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 874 ALKPSR 880
 |||||
 Db 42 alkpsr 48
 XX
 RESULT 11
 AAG03014
 ID AAG03014 standard; Protein; 61 AA.
 XX
 AC AAG03014;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein, SEQ ID NO: 7095.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 DR N-PSDB: AAC03020.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 13; SEQ ID 7095; 71pp + CD-ROM; English.
 XX
 CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or poly(A) RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs with genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX
 SQ Sequence 61 AA;
 XX
 Query Match 0.7%; Score 7; DB 21; Length 61;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 RETOSUD 992
 |||||
 Db 28 retsqid 34

RESULT 12

ID AAB39306 standard; Protein; 62 AA.

AC AAB39306;

DT 02-FEB-2001 (first entry)

DE Human secreted protein sequence encoded by gene 47 SEQ ID NO:186.

Human: secreted protein; immunosuppressive; antiarthritic; antirheumatic;
 antiapoptotic; cytosolic; cardiant; vasotropic; cerebroprotective;
 neurotrophic; neuroprotective; antibacterial; virucide; fungicide; neoplasm;
 optalmallogical; autoimmune disease; rheumatoid arthritis; angiogenesis;
 hyperproliferative disorder; cardiovascular disorder; infection;
 cerebrovascular disorder; nervous system disorder; ocular disorder;
 wound healing; chemotaxis.

OS Homo sapiens.

PN WO200056754-A1.

PD 28-SEP-2000.

PF 16-MAR-2000; 2000WO-US06792.

PR 19-MAR-1999; 99US-0125362.

PR 10-DEC-1999; 99US-0169980.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen GA, Ruben SM, Komatsoulis G;

DR WPI; 2000-579483/54.

DR N-PSDB; AAC74269.

PT Isolated nucleic acid molecule encoding a human secreted protein is
 used in preventing, treating or ameliorating a medical condition -

PS Disclosure; Page 69-70; 434pp; English.

The polynucleotide sequences given in AAC74223-C74279 encode the human
 secreted proteins represented in AAB39179-B39226. Sequences
 AAB39227-B39308 are alternative proteins encoded by the genes, and also
 protein sequences with which they share homology. The proteins have
 activities based on the tissues and cells in which they are expressed.
 Examples of activities include: immunosuppressive; antiarthritic;
 antirheumatic; antiapoptotic; cytosolic; cardiant; vasotropic;
 cerebroprotective; neurotrophic; neuroprotective; antibacterial; virucide;
 fungicide; and ophthalmallogical. The human secreted proteins,
 polynucleotides, antagonists and agonists of the invention may be useful
 in the treatment, prevention, and/or diagnosis of various disease,
 disorders and conditions such as autoimmune diseases e.g. rheumatoid
 arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
 liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
 disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders
 e.g. Alzheimer's disease, infections caused by bacteria, viruses and
 fungi and ocular disorders e.g. corneal infection. The polypeptides can
 also be used to aid wound healing and epithelial cell proliferation, to
 regenerate tissues, maintain organs before transplantation, in
 chemotaxis and as a food additive or preservative e.g. to increase
 storage capabilities. Sequences AAC74214-C74222 and AAB39178 are used
 during the isolation and characterisation of the genes of the invention.

Sequence 62 AA;

Query Match 0.7%; Score 7; DB 21; Length 62;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 669 N1YLSFH 675
 |||||
 Db 28 n1ylsfch 34

RESULT 13

ID ABB44131 standard; Peptide; 63 AA.

AC ABB44131;

DT 04-FEB-2002 (first entry)

DE Peptide #11637 encoded by human foetal liver single exon probe.

Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human fetal liver -

PS Claim 27; SEQ ID NO 36766; 639pp + sequence listing; English.

The invention relates to a single exon nucleic acid probe for
 measuring human gene expression in a sample derived from human foetal
 liver. The single exon nucleic acid probes may be used for predicting,
 measuring and displaying gene expression in samples derived from human
 fetal liver. The present sequence is a peptide encoded by a single exon
 nucleic acid probe of the invention.

Note: The sequence data for this patent did not form part of the
 printed specification, but was obtained in electronic format directly
 from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 63 AA;

Query Match 0.7%; Score 7; DB 22; Length 63;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SSRSSS 38
 |||||
 Db 27 ssrsgss 33

RESULT 14

ID ABB17718 standard; Protein; 63 AA.

PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-541565/60.
 DR N-PSDB; ABA14044.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Claim 11; SEQ ID NO 6375; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (ABA1678-ABA18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (anti)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 63 AA:
 Query Match 0.7%; Score 7; DB 22; Length 63;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 406 SLEVTY 412
 Db 13 SLEVTY 19
 RESULT 15
 ABB27013
 ID ABB27013 standard; Protein: 63 AA.
 XX
 AC ABB27013;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Protein #9012 encoded by probe for measuring heart cell gene expression.
 XX
 KW Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 XX
 OS Homo sapiens.

XX
 PN WO200157274-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00666.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488899/53.
 DR
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 XX
 PS Claim 15; SEQ ID NO 28783; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA1305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 63 AA:
 Query Match 0.7%; Score 7; DB 22; Length 63;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 32 SSRGSS 38
 Db 27 SSRGSS 33
 Search completed: May 21, 2002, 14:54:01
 Job time: 206 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 14:51:15 ; Search time 17.88 seconds
(without alignments)
1467.175 Million cell updates/sec

Title: US-09-729-653-2_COPY_1_1074
Perfect score: 1074
Sequence: 1 HSLIGRCSRGLGDNNAVAC.....LYFEGEKRYLQAGRFLLCG 1074

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	0.7	196	US-08-679-493A-185	Sequence 185, App
2	7	0.7	249	US-08-459-264-2	Sequence 2, Appl
3	7	0.7	249	US-08-459-264-3	Sequence 3, Appl
4	7	0.7	249	US-08-459-263-2	Sequence 2, Appl
5	7	0.7	249	US-08-459-263-3	Sequence 3, Appl
6	7	0.7	359	US-08-809-103B-2	Sequence 2, Appl
7	7	0.7	359	US-08-809-103B-4	Sequence 4, Appl
8	7	0.7	359	US-08-809-103B-6	Sequence 6, Appl
9	7	0.7	359	US-08-809-103B-8	Sequence 8, Appl
10	7	0.7	362	US-08-118-270-15	Sequence 15, Appl
11	7	0.7	362	PCT-US93-08528-15	Sequence 15, Appl
12	7	0.7	510	US-09-021-323-1	Sequence 1, Appl
13	7	0.7	574	US-08-907-166-12	Sequence 12, Appl
14	7	0.7	752	US-08-420-235B-21	Sequence 21, Appl
15	7	0.7	752	US-08-793-624-21	Sequence 21, Appl
16	7	0.7	752	PCT-US95-01194-21	Sequence 21, Appl
17	7	0.7	1612	US-08-169-927-2	Sequence 2, Appl
18	6	0.6	10	US-08-250-789A-67	Sequence 67, Appl
19	6	0.6	10	US-08-250-789A-76	Sequence 76, Appl
20	6	0.6	10	US-08-250-789A-83	Sequence 83, Appl
21	6	0.6	11	US-08-197-795-4	Sequence 4, Appl
22	6	0.6	11	PCT-US95-01721-4	Sequence 4, Appl
23	6	0.6	44	US-08-687-590-8	Sequence 8, Appl
24	6	0.6	44	US-08-687-590-9	Sequence 9, Appl
25	6	0.6	44	US-08-988-856B-6	Sequence 6, Appl
26	6	0.6	44	US-08-988-856B-16	Sequence 16, Appl
27	6	0.6	52	US-08-824-800D-12	Sequence 12, Appl

28	6	0.6	52	US-09-588-751-12	Sequence 12, Appl
29	6	0.6	56	US-08-867-087B-20	Sequence 20, Appl
30	6	0.6	64	US-09-627-376-6	Sequence 6, Appl
31	6	0.6	84	PCT-US93-08386-13	Sequence 13, Appl
32	6	0.6	94	US-08-466-368-5	Sequence 5, Appl
33	6	0.6	99	US-08-444-818-42	Sequence 42, Appl
34	6	0.6	100	US-08-672-850-13	Sequence 13, Appl
35	6	0.6	102	US-08-952-337-5	Sequence 5, Appl
36	6	0.6	102	US-08-894-173-75	Sequence 75, Appl
37	6	0.6	102	US-09-398-193-75	Sequence 75, Appl
38	6	0.6	103	US-08-472-171-2	Sequence 2, Appl
39	6	0.6	103	US-08-894-526-2	Sequence 2, Appl
40	6	0.6	103	US-09-013-047-2	Sequence 2, Appl
41	6	0.6	103	US-09-374-597-2	Sequence 2, Appl
42	6	0.6	103	US-08-857-076-35	Sequence 35, Appl
43	6	0.6	103	US-08-857-076-111	Sequence 111, Appl
44	6	0.6	104	US-09-012-515A-6	Sequence 6, Appl
45	6	0.6	104	US-08-360-144A-6	Sequence 6, Appl
46	6	0.6	104	PCT-US95-06722-6	Sequence 6, Appl
47	6	0.6	107	US-08-197-795-2	Sequence 2, Appl
48	6	0.6	107	US-08-336-618-13	Sequence 13, Appl
49	6	0.6	107	US-08-336-618-15	Sequence 15, Appl
50	6	0.6	107	US-08-963-601-1	Sequence 1, Appl
51	6	0.6	107	US-08-744-701-7	Sequence 7, Appl
52	6	0.6	107	US-08-894-173-67	Sequence 67, Appl
53	6	0.6	107	US-08-894-173-68	Sequence 68, Appl
54	6	0.6	107	US-08-894-173-80	Sequence 80, Appl
55	6	0.6	107	US-08-894-173-81	Sequence 81, Appl
56	6	0.6	107	US-09-398-193-67	Sequence 67, Appl
57	6	0.6	107	US-09-398-193-68	Sequence 68, Appl
58	6	0.6	107	US-09-398-193-80	Sequence 80, Appl
59	6	0.6	107	US-09-398-193-81	Sequence 81, Appl
60	6	0.6	107	PCT-US92-03993-2	Sequence 2, Appl
61	6	0.6	107	PCT-US92-03993-3	Sequence 3, Appl
62	6	0.6	107	PCT-US95-01721-2	Sequence 2, Appl
63	6	0.6	108	US-07-822-966B-5	Sequence 5, Appl
64	6	0.6	108	US-08-336-618-14	Sequence 14, Appl
65	6	0.6	108	US-08-803-899-3	Sequence 3, Appl
66	6	0.6	108	US-08-803-899-4	Sequence 4, Appl
67	6	0.6	108	US-08-961-083-166	Sequence 166, Appl
68	6	0.6	109	US-08-894-173-95	Sequence 95, Appl
69	6	0.6	109	US-09-398-193-95	Sequence 95, Appl
70	6	0.6	115	US-07-903-029-4	Sequence 4, Appl
71	6	0.6	115	US-07-903-029-5	Sequence 5, Appl
72	6	0.6	116	US-08-702-344-16	Sequence 16, Appl
73	6	0.6	117	US-08-444-818-44	Sequence 44, Appl
74	6	0.6	123	US-08-952-337-1	Sequence 1, Appl
75	6	0.6	124	US-08-449-045C-4	Sequence 4, Appl
76	6	0.6	124	US-08-435-605A-12	Sequence 12, Appl
77	6	0.6	124	US-09-012-097A-22	Sequence 22, Appl
78	6	0.6	128	US-08-687-816-20	Sequence 20, Appl
79	6	0.6	135	US-08-687-816-20	Sequence 20, Appl
80	6	0.6	135	US-09-138-614-20	Sequence 20, Appl
81	6	0.6	138	US-08-630-172-1	Sequence 1, Appl
82	6	0.6	138	US-09-375-419-1	Sequence 1, Appl
83	6	0.6	141	US-08-766-858A-46	Sequence 46, Appl
84	6	0.6	150	US-08-592-214A-14	Sequence 14, Appl
85	6	0.6	150	US-08-659-188-14	Sequence 14, Appl
86	6	0.6	150	US-08-659-227-14	Sequence 14, Appl
87	6	0.6	150	US-08-655-241-14	Sequence 14, Appl
88	6	0.6	150	US-09-149-976-14	Sequence 14, Appl
89	6	0.6	150	US-09-398-326-14	Sequence 14, Appl
90	6	0.6	152	US-08-672-850-6	Sequence 6, Appl
91	6	0.6	152	US-08-672-850-6	Sequence 6, Appl
92	6	0.6	152	US-08-672-850-12	Sequence 12, Appl
93	6	0.6	154	US-09-240-639-14	Sequence 14, Appl
94	6	0.6	159	US-08-844-086-4	Sequence 4, Appl
95	6	0.6	159	US-09-018-211-4	Sequence 4, Appl
96	6	0.6	160	US-08-961-539-4	Sequence 4, Appl
97	6	0.6	160	US-09-185-826-4	Sequence 4, Appl
98	6	0.6	165	US-09-335-409-16	Sequence 16, Appl
99	6	0.6	165	US-09-568-102-16	Sequence 16, Appl
100	6	0.6	185	US-09-567-969-16	Sequence 16, Appl

247	6	0.6	375	1	US-08-464-523B-24	Sequence 24, Appl	320	6	0.6	428	3	US-08-883-610A-2	Sequence 2, Appl
248	6	0.6	376	2	US-08-558-269-10	Sequence 10, Appl	321	6	0.6	428	4	US-08-936-094A-2	Sequence 2, Appl
249	6	0.6	376	4	US-09-410-882-10	Sequence 10, Appl	322	6	0.6	428	4	US-09-347-833-6	Sequence 6, Appl
250	6	0.6	377	1	US-08-188-277B-2	Sequence 2, Appl	323	6	0.6	429	2	US-08-310-912A-109	Sequence 109, App
251	6	0.6	377	1	US-08-188-277B-4	Sequence 4, Appl	324	6	0.6	429	2	US-09-301-085-109	Sequence 109, App
252	6	0.6	377	2	US-08-429-964-78	Sequence 78, Appl	325	6	0.6	429	5	PCT-US95-04589-109	Sequence 109, App
253	6	0.6	377	2	US-08-429-964-80	Sequence 80, Appl	326	6	0.6	432	3	US-08-477-460B-2	Sequence 2, Appl
254	6	0.6	378	1	US-08-707-793A-4	Sequence 4, Appl	327	6	0.6	432	3	US-08-379-516-2	Sequence 2, Appl
255	6	0.6	378	1	US-08-707-792A-4	Sequence 4, Appl	328	6	0.6	432	3	US-09-329-916-2	Sequence 2, Appl
256	6	0.6	380	2	US-08-846-762A-86	Sequence 86, Appl	329	6	0.6	432	4	US-08-485-972A-2	Sequence 2, Appl
257	6	0.6	383	1	US-08-464-523B-31	Sequence 31, Appl	330	6	0.6	432	4	US-09-409-006A-2	Sequence 2, Appl
258	6	0.6	383	2	US-08-558-269-6	Sequence 6, Appl	331	6	0.6	432	5	PCT-US93-07422-2	Sequence 2, Appl
259	6	0.6	383	4	US-09-410-882-6	Sequence 6, Appl	332	6	0.6	433	2	US-08-867-149-1	Sequence 1, Appl
260	6	0.6	384	1	US-08-707-793A-5	Sequence 5, Appl	333	6	0.6	433	2	US-08-808-374-1	Sequence 1, Appl
261	6	0.6	384	1	US-08-707-792A-5	Sequence 4, Appl	334	6	0.6	433	3	US-09-100-409A-1	Sequence 1, Appl
262	6	0.6	386	2	US-08-858-052-4	Sequence 4, Appl	335	6	0.6	433	4	US-09-400-208B-5	Sequence 5, Appl
263	6	0.6	386	3	US-08-200-284-4	Sequence 4, Appl	336	6	0.6	433	6	5171838-13	Patent No. 5171838
264	6	0.6	386	3	US-08-965-762-5	Sequence 5, Appl	337	6	0.6	434	1	US-08-236-311-4	Sequence 4, Appl
265	6	0.6	388	4	US-09-188-930-775	Sequence 275, App	338	6	0.6	434	1	US-08-162-392-4	Sequence 4, Appl
266	6	0.6	394	1	US-07-914-281-4	Sequence 4, Appl	339	6	0.6	434	1	US-08-643-300-4	Sequence 4, Appl
267	6	0.6	394	1	US-08-393-246-4	Sequence 4, Appl	340	6	0.6	434	1	US-08-643-297-4	Sequence 4, Appl
268	6	0.6	394	2	US-08-525-058A-4	Sequence 4, Appl	341	6	0.6	434	1	US-08-643-298-4	Sequence 4, Appl
269	6	0.6	394	2	US-08-636-731-4	Sequence 4, Appl	342	6	0.6	434	2	US-08-643-301-4	Sequence 4, Appl
270	6	0.6	394	3	US-08-466-368-2	Sequence 2, Appl	343	6	0.6	434	3	US-08-457-918-4	Sequence 4, Appl
271	6	0.6	394	4	US-09-042-531-4	Sequence 4, Appl	344	6	0.6	434	5	PCT-US94-14095-4	Sequence 4, Appl
272	6	0.6	394	5	PCT-US91-00899-9	Sequence 9, Appl	345	6	0.6	439	3	US-08-993-359-24	Sequence 24, Appl
273	6	0.6	394	6	5223418-2	Patent No. 5223418	346	6	0.6	439	3	US-09-321-654-2	Sequence 2, Appl
274	6	0.6	398	2	US-08-284-391B-29	Sequence 29, Appl	347	6	0.6	439	3	US-08-989-358A-2	Sequence 2, Appl
275	6	0.6	398	4	US-09-218-950-29	Sequence 29, Appl	348	6	0.6	439	4	US-09-457-046B-68	Sequence 68, Appl
276	6	0.6	399	4	US-08-765-907A-10	Sequence 10, Appl	349	6	0.6	441	4	US-09-171-661-26	Sequence 26, Appl
277	6	0.6	402	1	US-08-236-311-1	Sequence 1, Appl	350	6	0.6	443	3	US-09-120-365-69	Sequence 69, Appl
278	6	0.6	402	1	US-08-553-703A-1	Sequence 1, Appl	351	6	0.6	443	4	US-09-515-039-69	Sequence 69, Appl
279	6	0.6	402	1	US-08-553-703A-4	Sequence 4, Appl	352	6	0.6	448	3	US-09-039-555B-15	Sequence 15, Appl
280	6	0.6	402	2	US-09-006-021-1	Sequence 1, Appl	353	6	0.6	458	3	US-08-466-368-4	Sequence 4, Appl
281	6	0.6	402	2	US-09-006-021-4	Sequence 4, Appl	354	6	0.6	458	6	522394-7	Patent No. 522394
282	6	0.6	402	3	US-08-457-918-1	Sequence 1, Appl	355	6	0.6	461	1	US-07-796-361A-13	Sequence 13, Appl
283	6	0.6	404	1	US-08-203-716-2	Sequence 2, Appl	356	6	0.6	462	1	US-07-612-673-4	Sequence 4, Appl
284	6	0.6	404	1	US-08-242-663A-2	Sequence 2, Appl	357	6	0.6	462	1	US-08-539-666-4	Sequence 4, Appl
285	6	0.6	404	1	US-08-440-179-2	Sequence 2, Appl	358	6	0.6	462	2	US-08-417-495-5	Sequence 5, Appl
286	6	0.6	404	2	US-08-450-130A-1	Sequence 1, Appl	359	6	0.6	462	2	US-08-284-391B-5	Sequence 5, Appl
287	6	0.6	404	2	US-08-391-916A-2	Sequence 2, Appl	360	6	0.6	462	4	US-09-218-950-5	Sequence 5, Appl
288	6	0.6	404	2	US-08-573-890-2	Sequence 2, Appl	361	6	0.6	462	5	PCT-US92-01785-5	Sequence 5, Appl
289	6	0.6	404	3	US-08-394-189B-4	Sequence 4, Appl	362	6	0.6	462	5	PCT-US95-00454-5	Sequence 5, Appl
290	6	0.6	404	3	US-08-450-362A-1	Sequence 1, Appl	363	6	0.6	464	2	US-08-878-989-4	Sequence 4, Appl
291	6	0.6	404	3	US-08-258-287B-39	Sequence 39, Appl	364	6	0.6	464	3	US-09-056-783-2	Sequence 2, Appl
292	6	0.6	404	3	US-08-368-704C-39	Sequence 39, Appl	365	6	0.6	464	4	US-09-272-796-4	Sequence 4, Appl
293	6	0.6	404	3	US-08-954-536-18	Sequence 18, Appl	366	6	0.6	464	4	US-08-487-550-4	Sequence 4, Appl
294	6	0.6	404	4	US-09-039-657-2	Sequence 2, Appl	367	6	0.6	477	1	US-07-735-065-2	Sequence 2, Appl
295	6	0.6	404	4	US-08-748-547-2	Sequence 2, Appl	368	6	0.6	477	2	US-08-469-202-12	Sequence 12, Appl
296	6	0.6	404	4	US-08-908-436-3	Sequence 3, Appl	369	6	0.6	477	2	US-08-484-434C-12	Sequence 12, Appl
297	6	0.6	404	4	US-09-248-179-2	Sequence 2, Appl	370	6	0.6	480	2	US-08-570-227A-2	Sequence 2, Appl
298	6	0.6	404	4	US-09-069-023-30	Sequence 30, Appl	371	6	0.6	480	4	US-09-077-891-2	Sequence 2, Appl
299	6	0.6	404	5	PCT-US93-05705-4	Sequence 4, Appl	372	6	0.6	484	3	US-08-369-822C-21	Sequence 21, Appl
300	6	0.6	404	5	PCT-US95-06132-2	Sequence 2, Appl	373	6	0.6	484	3	US-08-582-776C-36	Sequence 36, Appl
301	6	0.6	404	5	PCT-US95-07619-1	Sequence 1, Appl	374	6	0.6	484	3	US-08-434-831B-33	Sequence 33, Appl
302	6	0.6	404	5	PCT-US96-10521-13	Sequence 13, Appl	375	6	0.6	491	2	US-08-687-916-23	Sequence 23, Appl
303	6	0.6	406	5	PCT-US94-02107-2	Sequence 2, Appl	376	6	0.6	491	2	US-09-001-826-23	Sequence 23, Appl
304	6	0.6	410	3	US-08-630-172-17	Sequence 17, Appl	377	6	0.6	491	4	US-09-138-614-23	Sequence 23, Appl
305	6	0.6	410	4	US-09-375-419-17	Sequence 17, Appl	378	6	0.6	492	2	US-08-644-271-32	Sequence 32, Appl
306	6	0.6	413	1	US-08-444-734A-5	Sequence 5, Appl	379	6	0.6	494	4	US-09-019-095A-24	Sequence 24, Appl
307	6	0.6	413	1	US-08-087-772A-17	Sequence 17, Appl	380	6	0.6	495	1	US-08-532-065B-2	Sequence 2, Appl
308	6	0.6	413	2	US-08-467-568-12	Sequence 12, Appl	381	6	0.6	495	2	US-08-687-916-22	Sequence 22, Appl
309	6	0.6	413	2	US-09-030-582-12	Sequence 12, Appl	382	6	0.6	495	4	US-09-138-614-22	Sequence 22, Appl
310	6	0.6	413	5	PCT-US91-00909-4	Sequence 4, Appl	383	6	0.6	496	2	US-08-949-637-2	Sequence 2, Appl
311	6	0.6	416	1	US-08-464-523B-33	Sequence 33, Appl	384	6	0.6	496	4	US-09-231-488-2	Sequence 2, Appl
312	6	0.6	417	1	US-08-464-523B-23	Sequence 23, Appl	385	6	0.6	499	4	US-08-988-856B-2	Sequence 2, Appl
313	6	0.6	418	4	US-09-198-603C-8	Sequence 8, Appl	386	6	0.6	499	4	US-08-988-856B-2	Sequence 2, Appl
314	6	0.6	423	1	US-08-464-523B-25	Sequence 25, Appl	387	6	0.6	499	5	PCT-US95-01806-2	Sequence 2, Appl
315	6	0.6	424	4	US-09-627-376-8	Sequence 8, Appl	388	6	0.6	499	5	PCT-US95-01806-4	Sequence 4, Appl
316	6	0.6	425	4	US-09-230-371A-30	Sequence 30, Appl	389	6	0.6	504	1	US-07-853-985A-6	Sequence 6, Appl
317	6	0.6	426	4	US-08-961-083-48	Sequence 48, Appl	390	6	0.6	504	1	US-07-681-703B-6	Sequence 6, Appl
318	6	0.6	427	2	US-09-196-857-2	Sequence 2, Appl	391	6	0.6	504	1	US-08-184-236-6	Sequence 6, Appl
319	6	0.6	428	2	US-08-717-587A-2	Sequence 2, Appl	392	6	0.6	504	2	US-08-407-410B-6	Sequence 6, Appl

685	6	0.6	836	5	PCT-US95-08071-105	758	6	0.6	1013	4	US-09-240-473-5	Sequence 5, Appl
686	6	0.6	845	6	5196194-17	759	6	0.6	1038	4	US-09-541-782-4	Sequence 4, Appl
687	6	0.6	846	1	US-08-356-354-2	760	6	0.6	1054	1	US-08-356-354-4	Sequence 4, Appl
688	6	0.6	846	2	US-07-728-215-33	761	6	0.6	1054	2	US-08-778-656-4	Sequence 4, Appl
689	6	0.6	846	2	US-08-778-656-2	762	6	0.6	1069	1	US-07-777-715-9	Sequence 9, Appl
690	6	0.6	846	4	US-08-938-085A-33	763	6	0.6	1069	1	US-08-170-126-4	Sequence 4, Appl
691	6	0.6	858	4	US-08-857-076-41	764	6	0.6	1069	1	US-08-954-418-4	Sequence 4, Appl
692	6	0.6	864	2	US-08-209-521-29	765	6	0.6	1084	3	US-08-717-515-6	Sequence 6, Appl
693	6	0.6	876	1	US-08-717-515-4	766	6	0.6	1090	3	US-08-307-896-3	Sequence 3, Appl
694	6	0.6	879	1	US-08-220-151-2	767	6	0.6	1090	3	US-08-726-214-4	Sequence 4, Appl
695	6	0.6	879	1	US-08-220-151-3	768	6	0.6	1090	4	US-09-085-199B-5	Sequence 5, Appl
696	6	0.6	879	1	US-08-413-118-2	769	6	0.6	1090	5	PCT-US95-11808-3	Sequence 3, Appl
697	6	0.6	879	1	US-08-413-118-3	770	6	0.6	1098	1	US-07-777-715-7	Sequence 7, Appl
698	6	0.6	879	3	US-08-413-118-106	771	6	0.6	1098	1	US-08-170-126-2	Sequence 2, Appl
699	6	0.6	879	3	US-08-473-446-2	772	6	0.6	1098	3	US-08-954-418-2	Sequence 2, Appl
700	6	0.6	879	3	US-08-473-446-3	773	6	0.6	1123	4	US-09-408-865-1	Sequence 1, Appl
701	6	0.6	879	3	US-08-473-446-106	774	6	0.6	1129	4	US-08-357-598-6	Sequence 6, Appl
702	6	0.6	892	4	US-08-857-076-42	775	6	0.6	1129	1	US-08-097-997A-9	Sequence 9, Appl
703	6	0.6	894	1	US-08-117-362-4	776	6	0.6	1129	2	US-08-567-508C-3	Sequence 3, Appl
704	6	0.6	894	1	US-08-486-924-4	777	6	0.6	1129	2	US-09-003-289-6	Sequence 6, Appl
705	6	0.6	898	2	US-08-808-982-5	778	6	0.6	1129	3	US-09-196-480-3	Sequence 9, Appl
706	6	0.6	898	4	US-09-306-902A-5	779	6	0.6	1129	4	US-08-665-574C-9	Sequence 9, Appl
707	6	0.6	903	1	US-08-021-601-12	780	6	0.6	1129	5	US-08-946-994-9	Sequence 9, Appl
708	6	0.6	903	1	US-08-082-849B-12	781	6	0.6	1132	2	PCT-US95-16435-6	Sequence 6, Appl
709	6	0.6	903	4	US-09-193-562D-46	782	6	0.6	1132	2	US-08-567-508C-2	Sequence 2, Appl
710	6	0.6	904	1	PCT-US94-01624-12	783	6	0.6	1132	3	US-09-196-480-2	Sequence 2, Appl
711	6	0.6	904	1	US-07-998-003A-97	784	6	0.6	1139	4	US-09-046-188A-22	Sequence 22, Appl
712	6	0.6	904	1	US-08-453-274B-97	785	6	0.6	1163	1	US-08-173-497-4	Sequence 4, Appl
713	6	0.6	904	1	US-08-453-695A-97	786	6	0.6	1163	1	US-08-286-889-4	Sequence 4, Appl
714	6	0.6	904	1	US-08-268-161A-97	787	6	0.6	1163	1	US-08-485-618-4	Sequence 4, Appl
715	6	0.6	904	2	US-08-453-702A-97	788	6	0.6	1163	1	US-08-362-652-4	Sequence 4, Appl
716	6	0.6	904	2	US-09-099-639-97	789	6	0.6	1163	2	US-08-605-672-4	Sequence 4, Appl
717	6	0.6	904	5	PCT-US93-12588-97	790	6	0.6	1163	2	US-08-482-293A-4	Sequence 4, Appl
718	6	0.6	905	4	PCT-US95-08071-97	791	6	0.6	1163	2	US-08-943-353-4	Sequence 4, Appl
719	6	0.6	905	4	US-09-193-562D-2	792	6	0.6	1163	2	US-08-476-062A-44	Sequence 44, Appl
720	6	0.6	910	4	US-08-460-269C-2	793	6	0.6	1163	2	US-09-193-043-4	Sequence 4, Appl
721	6	0.6	914	4	US-09-088-159B-4	794	6	0.6	1163	5	PCT-US96-01314-44	Sequence 44, Appl
722	6	0.6	921	1	US-08-188-582-2	795	6	0.6	1182	4	US-09-041-886-21	Sequence 21, Appl
723	6	0.6	921	1	US-08-646-715-2	796	6	0.6	1248	3	US-08-726-214-16	Sequence 16, Appl
724	6	0.6	924	3	US-08-619-812-8	797	6	0.6	1276	1	US-08-717-515-8	Sequence 8, Appl
725	6	0.6	926	1	US-07-908-253-2	798	6	0.6	1323	4	US-09-004-838-90	Sequence 90, Appl
726	6	0.6	926	1	US-08-455-970A-2	799	6	0.6	1334	6	US-08-425-061-23	Sequence 23, Appl
727	6	0.6	926	1	US-08-387-156-6	800	6	0.6	1363	1	US-08-825-886-23	Sequence 23, Appl
728	6	0.6	926	2	US-08-694-865-6	801	6	0.6	1382	3	US-09-057-570-4	Sequence 4, Appl
729	6	0.6	926	2	US-08-878-748-6	802	6	0.6	1403	3	US-07-908-253-3	Sequence 3, Appl
730	6	0.6	926	2	US-08-535-837-2	803	6	0.6	1403	2	US-08-694-865-17	Sequence 17, Appl
731	6	0.6	926	3	US-09-124-491-6	804	6	0.6	1403	2	US-08-535-837-3	Sequence 3, Appl
732	6	0.6	926	6	547657-3	805	6	0.6	1403	3	US-09-124-491-17	Sequence 17, Appl
733	6	0.6	928	3	US-09-320-878-13	806	6	0.6	1403	3	US-08-840-062-6	Sequence 6, Appl
734	6	0.6	928	4	US-09-105-537-41	807	6	0.6	1507	6	5268270-2	Sequence 6, Appl
735	6	0.6	934	1	US-08-215-805A-80	808	6	0.6	1507	6	US-08-825-886-23	Sequence 23, Appl
736	6	0.6	936	1	US-08-455-970A-12	809	6	0.6	1568	4	US-09-458-791-2	Sequence 2, Appl
737	6	0.6	943	1	US-08-455-970A-10	810	6	0.6	1568	4	US-09-459-066-2	Sequence 2, Appl
738	6	0.6	944	4	US-09-449-285A-2	811	6	0.6	1568	4	US-08-927-927-3	Sequence 3, Appl
739	6	0.6	951	4	US-08-455-970A-14	812	6	0.6	1621	4	US-09-057-570-2	Sequence 2, Appl
740	6	0.6	971	1	US-08-446-038B-19	813	6	0.6	1657	3	US-08-609-049A-13	Sequence 13, Appl
741	6	0.6	971	1	US-08-446-010B-19	814	6	0.6	1658	2	US-09-170-996-30	Sequence 30, Appl
742	6	0.6	971	2	US-08-805-445-19	815	6	0.6	1726	2	US-08-609-049A-30	Sequence 30, Appl
743	6	0.6	971	2	US-08-064-067D-19	816	6	0.6	1726	4	US-09-170-996-30	Sequence 30, Appl
744	6	0.6	971	2	US-09-066-208-19	817	6	0.6	1754	1	US-07-745-206A-13	Sequence 13, Appl
745	6	0.6	977	1	US-08-387-156-8	818	6	0.6	1754	1	US-08-444-818-54	Sequence 54, Appl
746	6	0.6	977	2	US-08-694-865-8	819	6	0.6	1786	4	US-08-311-363-13	Sequence 13, Appl
747	6	0.6	977	2	US-08-878-748-8	820	6	0.6	1788	2	US-08-962-284-2	Sequence 2, Appl
748	6	0.6	977	3	US-09-124-491-8	821	6	0.6	1792	2	US-08-962-284-4	Sequence 4, Appl
749	6	0.6	979	3	US-08-870-529-2	822	6	0.6	1805	3	US-09-057-570-7	Sequence 7, Appl
750	6	0.6	986	3	US-08-938-291A-7	823	6	0.6	1817	4	US-09-004-838-125	Sequence 125, App
751	6	0.6	993	4	US-08-446-010B-25	824	6	0.6	1817	2	US-08-680-327-3	Sequence 3, Appl
752	6	0.6	993	4	US-08-980-080-4	825	6	0.6	1824	2	US-09-228-246-2	Sequence 2, Appl
753	6	0.6	1000	2	US-09-193-562D-30	826	6	0.6	1852	1	US-08-425-061-24	Sequence 24, Appl
754	6	0.6	1003	2	US-07-743-357-9	827	6	0.6	1852	2	US-08-825-886-24	Sequence 24, Appl
755	6	0.6	1013	2	US-08-866-650-5	828	6	0.6	1863	1	US-08-425-061-16	Sequence 16, Appl
756	6	0.6	1013	2	US-09-021-287-5	829	6	0.6	1863	1	US-08-598-591-2	Sequence 2, Appl
757	6	0.6	1013	3	US-08-991-408-2	830	6	0.6	1863	1		


```

977 5 0.5 7 2 US-08-819-013-3 Sequence 3, Appli
978 5 0.5 7 3 US-08-474-410-38 Sequence 38, Appl
979 5 0.5 7 4 US-08-486-673B-38 Sequence 38, Appl
980 5 0.5 8 1 US-08-190-788A-19 Sequence 19, Appl
981 5 0.5 8 1 US-08-097-938-37 Sequence 37, Appl
982 5 0.5 8 1 US-08-476-000-37 Sequence 37, Appl
983 5 0.5 8 1 US-08-472-840-37 Sequence 37, Appl
984 5 0.5 8 1 US-08-383-474B-24 Sequence 24, Appl
985 5 0.5 8 1 US-08-465-391A-19 Sequence 19, Appl
986 5 0.5 8 2 US-08-464-538B-19 Sequence 19, Appl
987 5 0.5 8 2 US-08-476-976-37 Sequence 37, Appl
988 5 0.5 8 2 US-08-463-076E-63 Sequence 63, Appl
989 5 0.5 8 2 US-08-350-260A-436 Sequence 436, Appl
990 5 0.5 8 3 US-08-159-339A-1046 Sequence 1046, Ap
991 5 0.5 8 3 US-08-474-410-37 Sequence 37, Appl
992 5 0.5 8 4 US-08-486-673B-37 Sequence 37, Appl
993 5 0.5 9 1 US-08-288-405A-6 Sequence 6, Appl
994 5 0.5 9 1 US-08-097-938-36 Sequence 36, Appl
995 5 0.5 9 1 US-08-476-000-36 Sequence 36, Appl
996 5 0.5 9 1 US-08-472-840-36 Sequence 36, Appl
997 5 0.5 9 2 US-08-650-262-12 Sequence 12, Appl
998 5 0.5 9 2 US-08-476-976-36 Sequence 36, Appl
999 5 0.5 9 3 US-08-159-339A-742 Sequence 742, App
1000 5 0.5 9 3 US-08-159-339A-1054 Sequence 1054, Ap

```

ALIGNMENTS

```

RESULT 1
US-08-679-493A-185
; Sequence 185, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: TAYLOR, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 185
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Tetrahymena sp.
US-08-679-493A-185

```

```

Query Match 0.7%; Score 7; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1060 EKKYIOA 1066
Db 188 EKKYIOA 194

```

```

RESULT 2
US-08-459-264-2
; Sequence 2, Application US/08459264
; Patent No. 5705340
; GENERAL INFORMATION:
; APPLICANT: RASMUSSEN, Beth A
; APPLICANT: TALLY, Francis P
; APPLICANT: GUZMAN, Yakov
; TITLE OF INVENTION: DNA SEQUENCES AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: OF CLASS B BETA-LACTAMASE ENZYMES FROM BACTEROIDES
; TITLE OF INVENTION: FRAGILIS
; NUMBER OF SEQUENCES: 9

```

```

CORRESPONDENCE ADDRESS:
ADDRESS: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,264
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Robinson, Joseph R
REGISTRATION NUMBER: 33,448
REFERENCE/DOCKET NUMBER: 0646/1B026-US3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 249 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacteroides fragilis
IMMEDIATE SOURCE:
CLONE: corr amino acid
US-08-459-264-2

```

```

Query Match 0.7%; Score 7; DB 1; Length 249;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 619 DKVYTYV 625
Db 35 DKVYTYV 41

```

```

RESULT 3
US-08-459-264-3
; Sequence 3, Application US/08459264
; Patent No. 5705340
; GENERAL INFORMATION:
; APPLICANT: RASMUSSEN, Beth A
; APPLICANT: TALLY, Francis P
; APPLICANT: GUZMAN, Yakov
; TITLE OF INVENTION: DNA SEQUENCES AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: OF CLASS B BETA-LACTAMASE ENZYMES FROM BACTEROIDES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,264

```

FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Robinson, Joseph R
REGISTRATION NUMBER: 33,448
REFERENCE/DOCKET NUMBER: 0646/1B026-US3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 249 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacteroides fragilis
IMMEDIATE SOURCE:
CLONE: tal3635
US-08-459-264-3

Query Match 0.7%; Score 7; DB 1; Length 249;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 619 DKVTTYV 625
|||||
Db 35 DKVTTYV 41

RESULT 4
US-08-459-263-2
Sequence 2, Application US/08459263
Patent No. 5792642
GENERAL INFORMATION:
APPLICANT: RASMUSSEN, Beth A
APPLICANT: TALLY, Francis P
APPLICANT: GLUZMAN, Yakov
TITLE OF INVENTION: DNA SEQUENCES AND AMINO ACID SEQUENCES
TITLE OF INVENTION: OF CLASS B BETA-LACTAMASE ENZYMES FROM BACTEROIDES
TITLE OF INVENTION: FRAGILIS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,263
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Robinson, Joseph R
REGISTRATION NUMBER: 33,448
REFERENCE/DOCKET NUMBER: 0646/1B026-US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 249 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacteroides fragilis
IMMEDIATE SOURCE:
CLONE: ccrA amino acid
US-08-459-263-2

Query Match 0.7%; Score 7; DB 1; Length 249;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 619 DKVTTYV 625
|||||
Db 35 DKVTTYV 41

RESULT 5
US-08-459-263-3
Sequence 3, Application US/08459263
Patent No. 5792642
GENERAL INFORMATION:
APPLICANT: RASMUSSEN, Beth A
APPLICANT: TALLY, Francis P
APPLICANT: GLUZMAN, Yakov
TITLE OF INVENTION: DNA SEQUENCES AND AMINO ACID SEQUENCES
TITLE OF INVENTION: OF CLASS B BETA-LACTAMASE ENZYMES FROM BACTEROIDES
TITLE OF INVENTION: FRAGILIS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,263
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Robinson, Joseph R
REGISTRATION NUMBER: 33,448
REFERENCE/DOCKET NUMBER: 0646/1B026-US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 249 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacteroides fragilis
IMMEDIATE SOURCE:
CLONE: tal3635
US-08-459-263-3

Query Match 0.7%; Score 7; DB 1; Length 249;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 619 DKVYTV 625
DB 35 DKVYTV 41

RESULT 6

US-08-809-103B-2
; Sequence 2, Application US/08809103B
; Patent No. 6133505
; GENERAL INFORMATION:
; APPLICANT: GRONENBORN, Bruno
; TITLE OF INVENTION: PHYTOPATHOGENIC DNA VIRUS RESISTANT
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,103B
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94.11040
; FILING DATE: 15-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR95/01192
; FILING DATE: 15-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: USB94AL CNR TOM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 521-2297
; TELEFAX: (703) 685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-809-103B-2

Query Match 0.7%; Score 7; DB 4; Length 359;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 DKDGVVL 109
DB 105 DKDGVVL 111

RESULT 7

US-08-809-103B-4
; Sequence 4, Application US/08809103B
; Patent No. 6133505
; GENERAL INFORMATION:
; APPLICANT: GRONENBORN, Bruno
; TITLE OF INVENTION: PHYTOPATHOGENIC DNA VIRUS RESISTANT
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON

STREET: 745 South 23rd Street
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/809,103B
FILING DATE: 17-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 94.11040
FILING DATE: 15-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR95/01192
FILING DATE: 15-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: USB94AL CNR TOM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 521-2297
TELEFAX: (703) 685-0573
TELEX: 248425 EMBON

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-809-103B-4

Query Match 0.7%; Score 7; DB 4; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 DKDGVVL 109
DB 105 DKDGVVL 111

RESULT 8

US-08-809-103B-6
; Sequence 6, Application US/08809103B
; Patent No. 6133505
; GENERAL INFORMATION:
; APPLICANT: GRONENBORN, Bruno
; TITLE OF INVENTION: PHYTOPATHOGENIC DNA VIRUS RESISTANT
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,103B
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94.11040

FILING DATE: 15-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR95/01192
FILING DATE: 15-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: USB94AL CNR TOM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 521-2297
TELEFAX: (703) 685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-809-103B-6

Query Match 0.7%; Score 7; DB 4; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 DKDGDVL 109
|||||
DB 105 DKDGDVL 111

RESULT 9

US-08-809-103B-8
Sequence 8, Application US/08809103B
Patent No. 6133505
GENERAL INFORMATION:
APPLICANT: GROENBORN, Bruno
TITLE OF INVENTION: PHYTOPATHOGENIC DNA VIRUS RESISTANT
TITLE OF INVENTION: TRANSGENIC PLANTS AND SEEDS AND METHODS FOR OBTAINING SAME
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: YOUNG & THOMPSON
STREET: 745 South 23rd Street
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/809,103B
FILING DATE: 17-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 94.11040
FILING DATE: 15-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR95/01192
FILING DATE: 15-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: USB94AL CNR TOM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 521-2297
TELEFAX: (703) 685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-809-103B-8

Query Match 0.7%; Score 7; DB 4; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 DKDGDVL 109
|||||
DB 105 DKDGDVL 111

RESULT 10

US-08-118-270-15
Sequence 15, Application US/08118270
Patent No. 5508384
GENERAL INFORMATION:
APPLICANT: Murphy, Randall B.
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,270
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 362 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-118-270-15

Query Match 0.7%; Score 7; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 VGIWWSL 750
|||||
DB 2 VGIWWSL 8

RESULT 11

PCT-US93-08528-15
Sequence 15, Application PC/TUS9308528
GENERAL INFORMATION:

APPLICANT: New York University
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROADY AND NEWMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08528
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 362 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-08528-15

Query Match 0.7%; Score 7; DB 5; Length 362;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 VGIWMSL 750
DB 2 VGIWMSL 8

RESULT 12
US-09-021-323-1
Sequence 1, Application US/09021323
Patent No. 5929033
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Corley, Neil C.
APPLICANT: Yue, Henry
TITLE OF INVENTION: EXTRACELLULAR MUCOUS MATRIX
TITLE OF INVENTION: GLYCOPROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/021,323
FILING DATE: Filed Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0477 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SINTNOT13
CLONE: 1805538
US-09-021-323-1

Query Match 0.7%; Score 7; DB 2; Length 510;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SSRSGSS 38
DB 49 SSRSGSS 55

RESULT 13
US-08-907-166-12
Sequence 12, Application US/08907166
Patent No. 5948666
GENERAL INFORMATION:
APPLICANT: Callen, Walter
APPLICANT: Mathet, Eric
TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
FILE REFERENCE: 09010/027001
CURRENT FILING DATE: 1997-08-06
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 574
TYPE: PRT
ORGANISM: Aquifex pyrophillus
US-08-907-166-12

Query Match 0.7%; Score 7; DB 2; Length 574;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 SVVLGKK 249
DB 387 SVVLGKK 393

RESULT 14
US-08-420-235B-21
Sequence 21, Application US/08420235B
Patent No. 5801042
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 47

Job time: 184 sec

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/420,235B
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-420-235B-21

Query Match 0.7%: Score 7; DB 1; Length 752;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 RIFSLLE 49
|||||
Db 602 RIFSLLE 608

RESULT 15
US-08-793-624-21
; Sequence 21, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
; FILE REFERENCE: 45185-C-PCT-US/JPM
; CURRENT APPLICATION NUMBER: US/08/793,624C
; CURRENT FILING DATE: 1997-02-18
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 21
; LENGTH: 752
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
; US-08-793-624-21

Query Match 0.7%: Score 7; DB 4; Length 752;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 RIFSLLE 49
|||||
Db 602 RIFSLLE 608

Search completed: May 21, 2002, 14:54:19

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 14:52:01 ; Search time 26.02 Seconds
(without alignments)
3966.178 Million cell updates/sec

Title: US-09-729-653-2_COPY_1_1074
Perfect score: 1074
Sequence: 1 HSLIGRCSRGILGDNVAVC.....LYFEGEKRYLDQKFFLLCG 1074

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: PIR_71:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	0.8	519	2 T27880	hypothetical prote
2	8	0.7	176	2 T17821	hypothetical prote
3	8	0.7	241	2 A84283	hypothetical prote
4	8	0.7	316	2 AE3648	flagellar motor sw
5	8	0.7	394	1 KIBSGM	phosphoglycerate k
6	8	0.7	414	2 AC0575	hypothetical membr
7	8	0.7	519	2 E87233	probable secreted
8	8	0.7	642	2 T29967	hypothetical prote
9	8	0.7	843	2 T34618	NADH dehydrogenase
10	8	0.7	1214	2 T47659	spliceosomal-like
11	8	0.7	1238	2 T32625	hypothetical prote
12	8	0.7	1520	2 G98275	hypothetical prote
13	8	0.7	1520	2 AF3008	polyketide synthas
14	7	0.7	39	2 A96026	probable transpos
15	7	0.7	91	4 S14968	hypothetical NADH
16	7	0.7	94	2 F84497	En/spm-like transp
17	7	0.7	98	2 S26981	pvs protein 1 - Ki
18	7	0.7	100	2 D48223	ig heavy chain V r
19	7	0.7	110	2 A95296	hypothetical prote
20	7	0.7	123	2 AC6281	protein F10B6.16 l
21	7	0.7	130	2 AC2314	hypothetical prote
22	7	0.7	131	1 S30119	ribosomal protein
23	7	0.7	136	2 G64734	yacI protein - Esc
24	7	0.7	136	2 C90644	hypothetical prote
25	7	0.7	136	2 C85495	hypothetical prote
26	7	0.7	142	2 G87147	conserved hypotet
27	7	0.7	143	2 T40376	heat shock protein
28	7	0.7	144	1 R5EC15	ribosomal protein
29	7	0.7	144	2 F64094	ribosomal protein

30	7	0.7	144	2 F91149	50S ribosomal subu
31	7	0.7	144	2 B85995	50S ribosomal subu
32	7	0.7	144	2 AD1008	50S ribosomal chat
33	7	0.7	154	2 G86699	hypothetical prote
34	7	0.7	155	2 J10071	alpha-pilin - Mota
35	7	0.7	156	2 A41490	beta-fimbrial prote
36	7	0.7	157	2 A24434	hypothetical prote
37	7	0.7	157	2 AB2539	prepilin - Moraxel
38	7	0.7	158	2 A55851	myosin LC1 catalyt
39	7	0.7	160	2 A25571	fimbrial protein Q
40	7	0.7	164	2 A42460	FUN3 protein - ye
41	7	0.7	173	2 S23411	ubiquinol--cytochr
42	7	0.7	183	1 RDQFBR	superoxide dismuta
43	7	0.7	196	2 A39223	hypothetical prote
44	7	0.7	201	2 E70211	probable ISRM2011-
45	7	0.7	201	2 F95878	dihydrokaempferol
46	7	0.7	216	2 S61416	riboflavin-specific
47	7	0.7	219	1 F69500	glutathione transf
48	7	0.7	219	2 A26484	hypothetical prote
49	7	0.7	223	2 D70760	DNA topoisomerase
50	7	0.7	227	2 B48536	hypothetical prote
51	7	0.7	229	2 D82230	probable small nuc
52	7	0.7	232	2 C84706	hypothetical prote
53	7	0.7	248	2 C75140	beta-lactamase (EC
54	7	0.7	249	2 A35263	heme exporter prot
55	7	0.7	250	2 A82125	hypothetical prote
56	7	0.7	259	2 T29727	glutamate racemase
57	7	0.7	262	2 F84994	cger-4 protein - C
58	7	0.7	265	2 S19113	hypothetical prote
59	7	0.7	267	2 B64175	hypothetical 30.9
60	7	0.7	275	2 B64986	probable elongatio
61	7	0.7	275	2 G91011	hypothetical prote
62	7	0.7	275	2 A85856	hypothetical prote
63	7	0.7	282	2 B98166	hypothetical prote
64	7	0.7	282	2 A99205	hypothetical prote
65	7	0.7	282	2 AE3121	hypothetical prote
66	7	0.7	293	2 T47708	hypothetical prote
67	7	0.7	295	2 A49906	glucose-1-phosphat
68	7	0.7	302	1 RGEGB	melibiose operon r
69	7	0.7	302	2 D91266	regulator of melib
70	7	0.7	302	2 A86107	regulator of melib
71	7	0.7	305	2 JC4525	nucleic acid-bind
72	7	0.7	307	2 A10004	conserved membrane
73	7	0.7	310	2 H81161	cysteine synthase
74	7	0.7	310	2 A11022	melibiose operon r
75	7	0.7	312	2 E70376	exopolysphatase
76	7	0.7	314	2 H71802	probable outer mem
77	7	0.7	314	2 T31220	transposase homolo
78	7	0.7	315	2 C64716	rare lipoprotein A
79	7	0.7	315	2 T25297	hypothetical prote
80	7	0.7	330	2 T32080	hypothetical prote
81	7	0.7	332	2 S64375	probable tyrosine
82	7	0.7	333	2 AH0268	anthranilate phosp
83	7	0.7	337	1 DEJUGC	glyceraldenhyde-3-p
84	7	0.7	338	2 T47218	glyceraldenhyde-3-p
85	7	0.7	338	2 D87801	protein C10G11.7 l
86	7	0.7	340	2 E95844	probable sugar ABC
87	7	0.7	342	2 AB2397	ATP-binding prote
88	7	0.7	344	2 G70036	spore coat polysac
89	7	0.7	350	2 S47292	phenol 2-monooxyge
90	7	0.7	351	2 S50098	site-specific DNA-
91	7	0.7	352	2 T22464	hypothetical prote
92	7	0.7	353	1 S44308	phenol 2-monooxyge
93	7	0.7	353	2 S70629	Rad51 homolog mei-
94	7	0.7	356	2 A99901	D-alanine-D-alanin
95	7	0.7	359	2 S22593	hypothetical prote
96	7	0.7	359	2 S39211	gene C1 protein -
97	7	0.7	359	2 S39235	gene C1 protein -
98	7	0.7	364	1 MENZNC	matrix protein - N
99	7	0.7	364	1 MENZNV	matrix protein - N
100	7	0.7	365	1 B69114	conserved hypotet
101	7	0.7	370	1 W2WL39	E2 protein - human
102	7	0.7	371	2 A71683	hypothetical prote

103	7	0.7	379	2	G70918	hypothetical prote	176	7	0.7	906	2	G70767	probable helicase
104	7	0.7	392	2	D71175	hypothetical prote	177	7	0.7	943	2	F69543	ATP-dependent RNA
105	7	0.7	394	2	QJ1399	phosphoglycerate k	178	7	0.7	957	2	A49847	nitrite reductase
106	7	0.7	394	2	A72223	pantothenate metab	179	7	0.7	986	2	JC4825	glucan 1,4-beta-gl
107	7	0.7	396	2	E90238	threonine synthase	180	7	0.7	988	1	DJVZFP	DNA-directed DNA p
108	7	0.7	396	2	F75388	1-deoxy-D-xylulose	181	7	0.7	993	2	S48436	probable RNA helic
109	7	0.7	400	2	AD2322	phosphoglycerate k	182	7	0.7	1056	2	T02930	lysine-ketoglutarat
110	7	0.7	401	1	TWWTGY	phosphoglycerate k	183	7	0.7	1062	2	F89102	protein F25E5.1 [1
111	7	0.7	402	2	S75082	pantothenate metab	184	7	0.7	1063	2	S18211	hypothetical prote
112	7	0.7	406	2	T36632	probable oxidoredu	185	7	0.7	1099	2	T30307	rexS protein - Lac
113	7	0.7	415	2	T40535	probable arginine-	186	7	0.7	1151	2	T18297	zinc-finger protei
114	7	0.7	418	2	AC3519	isochorismate synt	187	7	0.7	1159	2	B98198	hypothetical prote
115	7	0.7	425	2	F70608	hypothetical prote	188	7	0.7	1159	2	AH3088	icmf (imported) -
116	7	0.7	433	2	D84667	probable PCI domai	189	7	0.7	1188	2	F64367	pyruvate, water dik
117	7	0.7	435	2	A72658	conserved Isocitrat	190	7	0.7	1297	2	S39791	neurotoxin - Clost
118	7	0.7	440	2	E71299	conserved hypothet	191	7	0.7	1334	2	T41524	rho1 gdp-gtp excha
119	7	0.7	450	2	S73419	signal recognition	192	7	0.7	1403	2	S24548	homeoic protein p
120	7	0.7	460	2	G90890	hypothetical prote	193	7	0.7	1474	2	T20488	hypothetical prote
121	7	0.7	460	2	A86180	hypothetical prote	194	7	0.7	1612	2	S59699	DNA topoisomerase
122	7	0.7	460	2	E84902	hypothetical prote	195	7	0.7	1626	2	A39242	outer membrane pro
123	7	0.7	460	2	B85727	hypothetical prote	196	7	0.7	1643	2	D71630	probable DNA-direc
124	7	0.7	462	2	T07989	acetyl-CoA C-acylt	197	7	0.7	1732	2	T43026	trichocyst protein
125	7	0.7	462	2	T52110	acetyl-CoA C-acylt	198	7	0.7	2090	2	S26058	t-complex polypept
126	7	0.7	468	2	A13241	conserved hypothet	199	7	0.7	2149	2	C96695	glucagon G1 - Nort
127	7	0.7	473	2	D81322	outer membrane pro	200	7	0.7	2391	2	G89779	glucagon G2 - Nort
128	7	0.7	481	2	B70908	hypothetical prote	201	7	0.7	2551	2	B98047	hypothetical prote
129	7	0.7	485	1	S32433	methyiaspartate mu	202	7	0.7	3305	2	T18358	apolipoprotein prec
130	7	0.7	488	2	F97039	hypothetical prote	203	6	0.6	15	2	S29174	D-galactose-bindin
131	7	0.7	489	2	T29924	hypothetical prote	204	6	0.6	28	2	A60698	trichocyst protein
132	7	0.7	498	2	T49747	hypothetical prote	205	6	0.6	29	2	A49410	t-complex polypept
133	7	0.7	510	2	A87482	conserved hypothet	206	6	0.6	31	2	S44471	glucagon G1 - Nort
134	7	0.7	515	2	F70904	hypothetical prote	207	6	0.6	31	2	S44472	glucagon G2 - Nort
135	7	0.7	517	2	T19962	hypothetical prote	208	6	0.6	31	2	C97398	hypothetical prote
136	7	0.7	520	2	D90014	hypothetical prote	209	6	0.6	32	2	S58524	alpha-complex prot
137	7	0.7	522	2	A83804	involved in spore	210	6	0.6	39	2	J00282	hypothetical 4.3K
138	7	0.7	531	2	T04463	hypothetical prote	211	6	0.6	44	2	F97562	hypothetical prote
139	7	0.7	535	2	B34576	D2 protein precurs	212	6	0.6	49	2	G85815	hypothetical prote
140	7	0.7	545	2	B82740	two-component syst	213	6	0.6	52	1	R3KM72	ribosomal protein
141	7	0.7	547	2	J00356	cellulase (EC 3.2.	214	6	0.6	53	2	T09190	hypothetical prote
142	7	0.7	551	2	S57447	HPBR11-7 protein -	215	6	0.6	53	2	T28821	hypothetical prote
143	7	0.7	566	2	S39732	arginine/ornithine	216	6	0.6	54	2	H90803	hypothetical prote
144	7	0.7	573	2	H86313	protein F2H15.10 [217	6	0.6	56	2	E69938	hypothetical prote
145	7	0.7	574	2	C70468	DNA polymerase I (218	6	0.6	57	2	S63680	signal transducer
146	7	0.7	576	2	A40688	peroxisomal protei	219	6	0.6	57	2	T10457	lipoprotein L - PS
147	7	0.7	589	2	H69691	two-component sens	220	6	0.6	61	2	A81898	hypothetical prote
148	7	0.7	591	2	A75474	serine proteinase,	221	6	0.6	63	2	S08190	metallothionein 1
149	7	0.7	595	2	T39681	transcription regu	222	6	0.6	64	2	T35034	hypothetical prote
150	7	0.7	596	2	A90534	hypothetical prote	223	6	0.6	65	2	D97706	hypothetical prote
151	7	0.7	605	2	D71053	hypothetical prote	224	6	0.6	66	2	S16201	photosystem I chai
152	7	0.7	607	2	A95898	probable sensor Ki	225	6	0.6	68	2	E70002	conserved hypothet
153	7	0.7	608	2	H75089	twlching mobility	226	6	0.6	68	2	F95222	hypothetical prote
154	7	0.7	612	2	I64241	glucose inhibited	227	6	0.6	68	2	F98086	hypothetical prote
155	7	0.7	625	2	T41603	alpha-amylase - fi	228	6	0.6	69	2	B91001	hypothetical prote
156	7	0.7	629	2	A95948	probable chaperoni	229	6	0.6	70	2	S35030	photosystem I chai
157	7	0.7	632	2	E81684	1-deoxyxylulose-5-	230	6	0.6	70	2	T06922	photosystem I chai
158	7	0.7	640	2	F71527	probable transkeo	231	6	0.6	71	2	I60082	CDA receptor - hum
159	7	0.7	654	2	G72344	phosphoglycerate k	232	6	0.6	71	2	H95299	hypothetical prote
160	7	0.7	664	2	AF1600	transketolase homo	233	6	0.6	72	2	E90412	hypothetical prote
161	7	0.7	678	2	T05821	hypothetical prote	234	6	0.6	76	2	G83935	thioredoxin relate
162	7	0.7	694	2	C72761	hypothetical prote	235	6	0.6	77	2	D82819	hypothetical prote
163	7	0.7	695	2	D71283	probable translati	236	6	0.6	78	2	AG2080	hypothetical prote
164	7	0.7	752	2	T16508	hypothetical prote	237	6	0.6	79	2	B64614	hypothetical prote
165	7	0.7	758	2	T16800	hypothetical prote	238	6	0.6	79	2	C71901	hypothetical prote
166	7	0.7	769	2	A71403	probable glucan sy	239	6	0.6	80	2	T28256	ORF MSV095 hypothe
167	7	0.7	784	2	E84785	probable protein k	240	6	0.6	81	2	S01410	hypothetical prote
168	7	0.7	803	2	H71437	probable protein k	241	6	0.6	82	2	D84562	hypothetical prote
169	7	0.7	807	2	T40821	probable ubiquitin	242	6	0.6	82	2	H83584	probable biotin-re
170	7	0.7	816	2	S54518	probable membrane	243	6	0.6	82	2	AD3480	hypothetical prote
171	7	0.7	825	2	H82885	hypothetical prote	244	6	0.6	84	2	A82916	ribosomal protein
172	7	0.7	846	2	C82135	chitinase VC1952 [245	6	0.6	84	2	C95923	hypothetical hypot
173	7	0.7	858	2	B86416	unknown protein, 1	246	6	0.6	85	2	T39205	dynamin light chain
174	7	0.7	858	2	B86416	unknown protein, 2	247	6	0.6	87	2	B69420	hypothetical prote
175	7	0.7	889	2	B55123	coatomer complex b	248	6	0.6	88	2	S68159	H+-transporting AT

249	6	0.6	88	2	C85866	hypothetical prote
250	6	0.6	89	2	F69858	hypothetical prote
251	6	0.6	89	2	AF3734	conserved hypotet
252	6	0.6	89	2	AG3394	hypothetical membr
253	6	0.6	90	2	C97737	hypothetical prote
254	6	0.6	91	2	H82370	conserved hypotet
255	6	0.6	91	2	B82709	hypothetical prote
256	6	0.6	92	2	S37507	Ig kappa chain V r
257	6	0.6	92	2	S21307	hypothetical prote
258	6	0.6	94	2	AG3478	hypothetical prote
259	6	0.6	95	2	D69121	hypothetical prote
260	6	0.6	97	2	H70197	stage V sporulatio
261	6	0.6	98	2	C91022	polymyxin resist
262	6	0.6	98	2	A64997	polymyxin B resist
263	6	0.6	98	2	D85866	polymyxin resist
264	6	0.6	98	2	A12262	hypothetical prote
265	6	0.6	99	2	E82930	conserved hypotet
266	6	0.6	100	2	E43550	hypothetical prote
267	6	0.6	101	2	C90075	intercellular adhe
268	6	0.6	101	2	D72318	hypothetical prote
269	6	0.6	102	2	S17687	rRNA N-glycosidase
270	6	0.6	102	2	S17688	rRNA N-glycosidase
271	6	0.6	102	2	S11894	hypothetical prote
272	6	0.6	102	2	S76695	hypothetical prote
273	6	0.6	103	2	B87261	hypothetical prote
274	6	0.6	104	2	AB1214	TN916 ORE33 homo
275	6	0.6	105	2	A48943	phase antigenic de
276	6	0.6	106	2	T49742	hypothetical prote
277	6	0.6	107	2	A61431	peptidylprolyl iso
278	6	0.6	107	2	T29406	hypothetical prote
279	6	0.6	107	2	T49594	hypothetical prote
280	6	0.6	108	1	K3HUB6	Ig kappa chain V-I
281	6	0.6	108	2	A35780	peptidylprolyl iso
282	6	0.6	108	2	UC5764	FK506-binding pro
283	6	0.6	108	2	A42657	FK506-binding pro
284	6	0.6	108	2	JH0528	FK506-binding pro
285	6	0.6	108	2	T26539	hypothetical prote
286	6	0.6	108	2	E30609	Ig kappa chain V-I
287	6	0.6	108	2	F64010	hypothetical prote
288	6	0.6	111	2	S74191	ribonuclease (EC 3
289	6	0.6	111	2	S11978	white protein VII -
290	6	0.6	111	2	H72707	hypothetical prote
291	6	0.6	112	2	AB3507	phosphoribosyl-Atp
292	6	0.6	112	2	D75451	hypothetical prote
293	6	0.6	114	2	T23119	hypothetical prote
294	6	0.6	114	2	G97938	hypothetical prote
295	6	0.6	115	1	A46279	guanylin precursor
296	6	0.6	115	2	S03243	hypothetical prote
297	6	0.6	116	1	S20630	somatosatin precu
298	6	0.6	116	2	T49862	related to REM1 pr
299	6	0.6	117	2	S34246	hypothetical prote
300	6	0.6	117	2	T17315	hypothetical prote
301	6	0.6	119	2	T19321	hypothetical prote
302	6	0.6	119	2	T29306	hypothetical prote
303	6	0.6	119	2	A85816	unknown protein en
304	6	0.6	119	2	B75331	conserved hypotet
305	6	0.6	120	2	C91027	D-erythro-7,8-dihy
306	6	0.6	120	2	D85871	D-erythro-7,8-dihy
307	6	0.6	120	2	E65002	D-erythro-7,8-dihy
308	6	0.6	120	2	F81701	hypothetical prote
309	6	0.6	120	2	H97515	hypothetical prote
310	6	0.6	121	2	T02926	acyl carrier prote
311	6	0.6	121	2	G64315	hypothetical prote
312	6	0.6	122	2	D83448	succinate dehydrog
313	6	0.6	122	2	E83111	transcription fact
314	6	0.6	122	2	A41810	hypothetical prote
315	6	0.6	122	2	B75387	hypothetical prote
316	6	0.6	123	2	H90672	hypothetical prote
317	6	0.6	123	2	E84752	hypothetical prote
318	6	0.6	124	1	COBY26	cell division cont
319	6	0.6	124	1	XVVCB	cholera enterotoxi
320	6	0.6	124	1	F72615	hypothetical prote
321	6	0.6	124	2	B82444	hypothetical prote
322	6	0.6	125	2	T03283	hypothetical prote
323	6	0.6	125	2	S23541	basic leucine zipp
324	6	0.6	125	2	UC4799	conserved hypotet
325	6	0.6	125	2	AE1047	ribosomal protein
326	6	0.6	126	2	T40732	probable 50s ribos
327	6	0.6	126	2	C70307	conserved hypotet
328	6	0.6	126	2	AB2905	hypothetical prote
329	6	0.6	126	2	AG1932	hypothetical prote
330	6	0.6	126	2	T47476	hypothetical prote
331	6	0.6	127	2	PC2266	cytochrome P450 pr
332	6	0.6	127	2	T20707	hypothetical prote
333	6	0.6	127	2	H64012	sufti protein homol
334	6	0.6	128	2	S27919	hypothetical prote
335	6	0.6	128	2	J02182	hypothetical 14.8k
336	6	0.6	128	2	G81035	hypothetical prote
337	6	0.6	128	2	D64251	RNaseP C5 chain -
338	6	0.6	128	2	AD1648	hypothetical prote
339	6	0.6	129	2	C69488	LSU ribosomal prot
340	6	0.6	129	2	S63743	pol polyprotein -
341	6	0.6	129	2	D90367	hypothetical prote
342	6	0.6	129	2	G90022	hypothetical prote
343	6	0.6	130	2	G70799	hypothetical prote
344	6	0.6	130	2	AD2901	hypothetical prote
345	6	0.6	131	1	RDEB15	hypothetical prote
346	6	0.6	131	2	S14390	hypothetical prote
347	6	0.6	131	2	J02171	conserved hypotet
348	6	0.6	131	2	S27911	hypothetical prote
349	6	0.6	131	2	AB1525	hypothetical prote
350	6	0.6	132	1	F71177	hypothetical prote
351	6	0.6	132	2	E75164	hypothetical prote
352	6	0.6	132	2	B71954	hypothetical prote
353	6	0.6	132	2	B64554	conserved hypotet
354	6	0.6	132	2	T46905	hypothetical prote
355	6	0.6	133	2	G72537	hypothetical prote
356	6	0.6	133	2	S58123	chlorodoxin (clone
357	6	0.6	133	2	T08447	hypothetical prote
358	6	0.6	133	2	T30655	hypothetical prote
359	6	0.6	133	2	D97680	hypothetical prote
360	6	0.6	133	2	A41328	hypothetical prote
361	6	0.6	133	2	G70727	hypothetical prote
362	6	0.6	134	1	G65129	cytidine deaminase
363	6	0.6	134	2	AD2592	probable regulator
364	6	0.6	134	2	G95280	hypothetical prote
365	6	0.6	134	2	G91154	hypothetical prote
366	6	0.6	134	2	D86000	hypothetical prote
367	6	0.6	134	2	AB1002	conserved hypotet
368	6	0.6	134	2	AC1166	hypothetical prote
369	6	0.6	134	2	H90235	hypothetical prote
370	6	0.6	134	2	B86199	hypothetical prote
371	6	0.6	135	2	S00388	T-cell receptor ga
372	6	0.6	135	2	AB3150	hypothetical prote
373	6	0.6	135	2	AC1447	Porlein gpII (Bact
374	6	0.6	135	2	AD1361	hypothetical prote
375	6	0.6	136	2	G97606	hypothetical prote
376	6	0.6	136	2	A12828	conserved hypotet
377	6	0.6	137	2	E69359	hypothetical prote
378	6	0.6	137	2	C57600	DNA-binding protel
379	6	0.6	138	2	AC3337	hypothetical prote
380	6	0.6	139	2	H72016	ribonuclease p pro
381	6	0.6	139	2	D86607	4-carboxymuolacto
382	6	0.6	139	2	G90466	hypothetical prote
383	6	0.6	139	2	C87459	hypothetical prote
384	6	0.6	139	2	PC4217	hypothetical 139 p
385	6	0.6	139	2	B95304	hypothetical prote
386	6	0.6	140	2	AB2127	hypothetical prote
387	6	0.6	140	2	F72263	flagellar basal-bo
388	6	0.6	140	2	T10925	3C3.16c protein -
389	6	0.6	140	2	T46220	hypothetical prote
390	6	0.6	140	2	A11741	protein involved i
391	6	0.6	140	2	G86413	F28W24.17 protein
392	6	0.6	141	1	HATPI	hemoglobin alpha-I
393	6	0.6	141	1	HATP2B	hemoglobin alpha-I
394	6	0.6	141	1	H69385	hypothetical prote

395	6	0.6	141	2	S55247	hemoglobin alpha-A
396	6	0.6	141	2	D64530	ribosomal protein
397	6	0.6	141	2	H71975	ribosomal protein
398	6	0.6	141	2	T40501	60s ribosomal prot
399	6	0.6	141	2	D69857	conserved hypotnet
400	6	0.6	141	2	D82939	small protein B uu
401	6	0.6	141	2	E70869	hypothetical prote
402	6	0.6	141	2	H64361	hypothetical prote
403	6	0.6	141	2	AC1372	proteins involved
404	6	0.6	142	2	JH0586	ribosomal protein
405	6	0.6	142	2	B70634	probable mmpsi pro
406	6	0.6	142	2	A95132	hypothetical prote
407	6	0.6	142	2	AP1808	hypothetical prote
408	6	0.6	143	2	G96072	7,8-dihydro-8-oxog
409	6	0.6	143	2	S73515	pts system manito
410	6	0.6	144	2	F84988	50S ribosomal prot
411	6	0.6	144	2	S43766	ribosomal protein
412	6	0.6	144	2	T43774	ribosomal protein
413	6	0.6	144	2	T64121	DNA-directed DNA p
414	6	0.6	144	2	A82263	hypothetical prote
415	6	0.6	144	2	AE3423	hypothetical prote
416	6	0.6	144	2	G69255	hypothetical prote
417	6	0.6	144	2	D96967	probable metal-dep
418	6	0.6	145	2	F97374	Cyldine deaminase
419	6	0.6	145	2	A87618	C factor limported
420	6	0.6	145	2	E84938	flagellar filii pro
421	6	0.6	146	1	G69110	conserved hypotnet
422	6	0.6	146	2	F70313	hypothetical prote
423	6	0.6	146	2	T48566	hemoglobin epsilon
424	6	0.6	147	1	HECH	ribosomal protein
425	6	0.6	147	2	T35563	hypothetical prote
426	6	0.6	147	2	D98205	hypothetical prote
427	6	0.6	147	2	AD3081	conserved hypotnet
428	6	0.6	147	2	C95227	conserved hypotnet
429	6	0.6	147	2	G98091	conserved hypotnet
430	6	0.6	147	2	D69544	hypothetical prote
431	6	0.6	147	2	S50775	hypothetical prote
432	6	0.6	147	2	H82172	heat shock protein
433	6	0.6	147	2	I40475	hypothetical prote
434	6	0.6	147	2	S72994	hypothetical prote
435	6	0.6	148	2	AE0928	probable phage tai
436	6	0.6	148	2	A11058	response membrane
437	6	0.6	149	2	F75609	response regulator
438	6	0.6	149	2	T37099	hypothetical prote
439	6	0.6	150	2	T14457	MAUS box protein h
440	6	0.6	150	2	H83632	conserved hypotnet
441	6	0.6	151	1	S73506	PIIB homolog K05.o
442	6	0.6	151	2	I38365	epididymal secreto
443	6	0.6	151	2	I53929	epididymal secreto
444	6	0.6	152	2	G95958	probable transcrip
445	6	0.6	153	1	WRBP15	early protein gp17
446	6	0.6	153	2	AF1334	hypothetical prote
447	6	0.6	154	2	A44809	fimbrial protein -
448	6	0.6	154	2	D82723	transcription elon
449	6	0.6	155	2	S32194	ribosomal protein
450	6	0.6	155	2	T49854	hypothetical prote
451	6	0.6	155	2	D87414	hypothetical prote
452	6	0.6	156	2	G82952	ATP synthase B cha
453	6	0.6	156	2	T43957	hypothetical prote
454	6	0.6	156	2	C97097	probable membrane
455	6	0.6	157	2	H64376	phosphoribosylamin
456	6	0.6	157	2	S17934	rRNA N-glycosidase
457	6	0.6	157	2	T10732	intracellular path
458	6	0.6	158	2	T12958	copper transport p
459	6	0.6	158	2	D85523	hypothetical prote
460	6	0.6	159	2	A47699	type 4 pilin - Bix
461	6	0.6	159	2	G82924	hypothetical uun80
462	6	0.6	159	2	H81657	conserved hypotnet
463	6	0.6	159	2	B84052	hypothetical prote
464	6	0.6	160	2	S40063	fimbrial protein c
465	6	0.6	160	2	S74317	fimbrial protein c
466	6	0.6	160	2	A81562	molymbdenum cofacto
467	6	0.6	160	2	AF1205	molymbdenum cofacto
468	6	0.6	160	2	C70947	hypothetical prote
469	6	0.6	160	2	AH2342	hypothetical prote
470	6	0.6	161	2	T14817	pathogenesis-relat
471	6	0.6	161	2	D97145	molymbdenum cofacto
472	6	0.6	162	1	S07661	probable phosphot
473	6	0.6	162	2	C33739	hypothetical 17.8k
474	6	0.6	163	2	T11552	thioredoxin peroxi
475	6	0.6	163	2	B43583	conserved hypotnet
476	6	0.6	163	2	D75504	hypothetical prote
477	6	0.6	164	2	T35813	hypothetical prote
478	6	0.6	165	2	S35195	molymbdenum cofacto
479	6	0.6	165	2	G84027	probable alanine-r
480	6	0.6	165	2	T36924	probable transcrip
481	6	0.6	165	2	T52114	retroviral protein
482	6	0.6	165	2	J50065	sucaine dehydrog
483	6	0.6	166	2	F97676	hycoaxanthine phosp
484	6	0.6	166	2	H84952	early protein gp17
485	6	0.6	166	2	JN0030	chemotaxis protein
486	6	0.6	166	2	H87652	NBS-LRR type resis
487	6	0.6	166	2	T02222	signal peptidase (
488	6	0.6	167	1	S48484	caltractin-like pr
489	6	0.6	167	2	B85437	flm3 region hypot
490	6	0.6	167	2	B53293	hypothetical prote
491	6	0.6	167	2	SA0978	hypothetical prote
492	6	0.6	168	2	H80399	phospholipid-hydro
493	6	0.6	169	2	T04207	hypothetical prote
494	6	0.6	169	2	F72465	adenine phosphorib
495	6	0.6	170	2	B99552	molymbdenum cofacto
496	6	0.6	170	2	H69786	conserved hypotnet
497	6	0.6	170	2	G69541	diol dehydrase (di
498	6	0.6	170	2	AC1219	diol dehydrase (di
499	6	0.6	170	2	AF1572	hypothetical prote
500	6	0.6	170	2	T48974	probable cyldine/
501	6	0.6	171	2	G87218	hypothetical prote
502	6	0.6	171	2	E83684	hypothetical prote
503	6	0.6	171	2	S15911	hypothetical prote
504	6	0.6	172	2	T32120	hypothetical prote
505	6	0.6	173	2	H82149	crossover junction
506	6	0.6	174	1	WRBP70	early protein gp17
507	6	0.6	174	2	E75552	ribosomal protein
508	6	0.6	174	2	T01649	probable trypsin 1
509	6	0.6	176	2	E90160	conserved hypotnet
510	6	0.6	176	2	C68824	hypothetical prote
511	6	0.6	176	2	C55208	socA3 protein - My
512	6	0.6	177	2	S36914	ribosomal protein
513	6	0.6	177	2	T37444	probable 20.7k pro
514	6	0.6	177	2	T40970	very hypothetical
515	6	0.6	178	2	S39875	hypothetical prote
516	6	0.6	178	2	G83844	hypothetical prote
517	6	0.6	178	2	T26853	hypothetical prote
518	6	0.6	179	2	AD1391	H+-transporting AT
519	6	0.6	179	2	B84008	hypothetical prote
520	6	0.6	179	2	AB1850	hypothetical prote
521	6	0.6	179	2	AF1766	H+-transporting AT
522	6	0.6	180	2	H81343	molymbdopterin bios
523	6	0.6	180	2	T18161	hypothetical prote
524	6	0.6	180	2	J02394	W protein - Newcas
525	6	0.6	180	2	A60670	aducanin - mouse (i
526	6	0.6	180	2	T13878	matuase-like prot
527	6	0.6	181	2	T04398	hypothetical prote
528	6	0.6	181	2	AC2124	hypothetical prote
529	6	0.6	182	2	J01801	B7R 21.3k protein
530	6	0.6	182	2	G75343	conserved hypotnet
531	6	0.6	183	1	C70008	pyrazinamide/nic
532	6	0.6	183	1	S42547	glucose-1-phosphat
533	6	0.6	183	2	G89872	conserved hypotnet
534	6	0.6	184	2	S75762	hypothetical prote
535	6	0.6	184	2	A37201	aminoglycoside N6'
536	6	0.6	184	2	E71002	hypothetical prote
537	6	0.6	185	2	JK0021	somatotropin - ski
538	6	0.6	185	2	A97118	probable sigma fac
539	6	0.6	186	2	JC2400	PMS5 homolog misma
540	6	0.6	187	2	PC1082	somatotropin - Jap

541	6	0.6	187	2	JU0030	somatotropin - yel	614	6	0.6	206	1	I64088	conserved hypotet
542	6	0.6	187	2	PC1088	somatotropin - gre	615	6	0.6	206	2	F87175	pyridoxamine 5'-ph
543	6	0.6	187	2	PC1087	somatotropin - duc	616	6	0.6	206	4	G47070	hypothetical tyvel
544	6	0.6	187	2	F64834	fimbrial-like prot	617	6	0.6	207	2	B70447	riboflavin synthas
545	6	0.6	187	2	C90757	probable fimbrial-	618	6	0.6	207	2	H84273	flagellin A1 precu
546	6	0.6	187	2	AC2190	polypeptide deform	619	6	0.6	207	2	F83815	hypothetical prote
547	6	0.6	188	2	F64496	GMP synthetase (EC	620	6	0.6	207	2	D86225	hypothetical prote
548	6	0.6	188	2	C87216	probable hemagglut	621	6	0.6	207	2	F83148	hypothetical prote
549	6	0.6	188	2	AB4214	adenine phosphorib	622	6	0.6	207	2	T39349	probable 60S ribos
550	6	0.6	188	2	T15651	hypothetical prote	623	6	0.6	209	2	A71313	probable 60S ribos
551	6	0.6	189	2	T43766	hypothetical prote	624	6	0.6	209	2	S16864	gene E161 protein
552	6	0.6	190	2	AH0156	probable elongatio	625	6	0.6	209	2	AF2037	hypothetical prote
553	6	0.6	191	2	S09635	paba protein - Ser	626	6	0.6	209	2	B81809	hypothetical prote
554	6	0.6	191	2	AF0021	para-aminobenzoate	627	6	0.6	209	2	G83359	conserved hypotet
555	6	0.6	191	2	T07994	ribosomal protein	628	6	0.6	210	2	B84775	probable harpin-in
556	6	0.6	191	2	E64909	probable phage-rel	629	6	0.6	210	2	C97157	stage III sporulat
557	6	0.6	191	2	H64887	probable phage-rel	630	6	0.6	210	2	S64376	hypothetical prote
558	6	0.6	192	1	E64087	lipoprotein B - Ha	631	6	0.6	210	2	E84499	hypothetical prote
559	6	0.6	193	2	G64187	anthranilate synth	632	6	0.6	210	2	E87395	hypothetical prote
560	6	0.6	193	2	AD1949	hypothetical prote	633	6	0.6	210	2	S67771	endoplasmic reticu
561	6	0.6	193	2	F81779	probable transpos	634	6	0.6	211	2	E70045	two-component resp
562	6	0.6	194	1	QX8P3L	hypothetical prote	635	6	0.6	211	2	C83360	hypothetical prote
563	6	0.6	194	1	G64026	acyl-carrier-prot	636	6	0.6	212	2	AB1190	conserved hypotet
564	6	0.6	195	2	E27733	nifH protein - Azo	637	6	0.6	213	2	E95232	ABC transporter, A
565	6	0.6	195	2	A90467	hypothetical prote	638	6	0.6	213	2	G98096	hypothetical prote
566	6	0.6	196	1	HSXLSA	histone H5B - Afri	639	6	0.6	214	2	T45714	endochitinase-like
567	6	0.6	196	2	A30484	histone H5B - Afri	640	6	0.6	214	2	G83488	probable permease
568	6	0.6	196	2	S06555	finger protein (cl	641	6	0.6	214	2	C71120	hypothetical prote
569	6	0.6	197	1	NMPS2P	anthranilate synth	642	6	0.6	214	2	C72372	pyrazinamidase/nic
570	6	0.6	197	2	T50025	hypothetical prote	643	6	0.6	214	2	S57813	hypothetical prote
571	6	0.6	197	2	C86748	hypothetical prote	644	6	0.6	214	2	T27259	hypothetical prote
572	6	0.6	197	2	H86741	conserved hypotet	645	6	0.6	214	2	A90028	urase accessory p
573	6	0.6	197	2	H57145	hypothetical prote	646	6	0.6	215	1	D71038	hypothetical prote
574	6	0.6	198	2	AB0112	probable 5-formyl	647	6	0.6	215	2	E96493	hypothetical prote
575	6	0.6	198	2	B89939	hypothetical prote	648	6	0.6	215	2	AB3023	probable transcrip
576	6	0.6	198	2	C70570	hypothetical prote	649	6	0.6	216	2	T35881	thymidine kinase (
577	6	0.6	198	2	T13215	hypothetical prote	650	6	0.6	216	2	JE0297	DRE/CRT-binding pr
578	6	0.6	198	2	B84870	probable molybdopt	651	6	0.6	216	2	T51830	transcription fact
579	6	0.6	199	2	F70742	hypothetical prote	652	6	0.6	216	2	S48486	probable membrane
580	6	0.6	199	2	C75213	hypothetical prote	653	6	0.6	217	2	T42053	red2 protein - Str
581	6	0.6	200	1	H70409	conserved hypotet	654	6	0.6	217	2	F72293	5-methylthioadenos
582	6	0.6	200	2	A87671	cytochrome c oxida	655	6	0.6	217	2	A72024	conserved hypotet
583	6	0.6	200	2	T35699	probable sigma fac	656	6	0.6	217	2	A86600	CT734 hypothetical
584	6	0.6	200	2	T08966	hypothetical prote	657	6	0.6	217	2	G84380	hypothetical prote
585	6	0.6	200	2	B90300	hypothetical prote	658	6	0.6	217	2	T14778	hypothetical prote
586	6	0.6	200	2	AF1495	probable sugar-pho	659	6	0.6	218	2	JC1520	site-specific DNA
587	6	0.6	201	2	B88431	protein M88.2 (imp	660	6	0.6	218	2	S75269	signal peptidase I
588	6	0.6	201	2	A87680	RNA polymerase sig	661	6	0.6	218	2	G82786	conserved hypotet
589	6	0.6	201	2	B72739	hypothetical prote	662	6	0.6	218	2	B82804	hypothetical prote
590	6	0.6	201	2	A13328	hypothetical prote	663	6	0.6	218	2	G72754	probable phosphoen
591	6	0.6	202	2	AG3507	imidazoleglycerol-	664	6	0.6	219	2	T45352	pyridoxamine 5'-ph
592	6	0.6	202	2	G84502	hypothetical prote	665	6	0.6	219	2	S72721	amidotransferase h
593	6	0.6	202	2	AB1441	probable orotate p	666	6	0.6	219	2	AF0639	flagellar basal bo
594	6	0.6	202	2	F97342	transcription regu	667	6	0.6	219	2	T38129	c-8 sterol isomera
595	6	0.6	202	2	E95883	probable transcrip	668	6	0.6	220	1	ISPUCH	chalcone isomerase
596	6	0.6	203	2	S00747	somatotropin precu	669	6	0.6	220	1	S01765	GTP-binding protei
597	6	0.6	203	2	C71375	conserved hypotet	670	6	0.6	220	2	D70350	dethiobiotin synth
598	6	0.6	203	2	H70336	hypothetical prote	671	6	0.6	220	2	H69257	hypothetical prote
599	6	0.6	204	1	STFL	somatotropin precu	672	6	0.6	221	2	T35154	hypothetical prote
600	6	0.6	204	2	S22616	hypothetical prote	673	6	0.6	221	2	S73349	adhesin p1 precurs
601	6	0.6	204	2	S01746	somatotropin precu	674	6	0.6	221	2	AF3195	transcription regu
602	6	0.6	204	2	JH0577	somatotropin precu	675	6	0.6	222	1	R5MX1	ribosomal protein
603	6	0.6	204	2	S30491	somatotropin - Aca	676	6	0.6	222	2	B64895	hypothetical prote
604	6	0.6	204	2	A56904	somatotropin precu	677	6	0.6	222	2	D90883	hypothetical prote
605	6	0.6	204	2	JC4261	somatotropin precu	678	6	0.6	222	2	C85735	hypothetical prote
606	6	0.6	204	2	AF0914	conserved hypotet	679	6	0.6	222	2	AE0668	probable lipoprote
607	6	0.6	204	2	T03962	r40g3 protein - ri	680	6	0.6	222	2	S77526	hypothetical prote
608	6	0.6	204	2	S67295	probable membrane	681	6	0.6	222	2	C75539	conserved hypotet
609	6	0.6	205	2	S30739	hypothetical prote	682	6	0.6	223	1	S72693	dethiobiotin synth
610	6	0.6	205	2	H91225	hypothetical prote	683	6	0.6	223	1	YLHNP	serum amyloid P-co
611	6	0.6	205	2	H86072	hypothetical prote	684	6	0.6	223	2	B86968	conserved hypotet
612	6	0.6	205	2	H83689	hypothetical prote	685	6	0.6	223	2	T26763	hypothetical prote
613	6	0.6	205	2	A97161	cell division prot	686	6	0.6	223	2	T35908	probable membrane

687	6	0.6	223	2	S06576	760	6	0.6	239	2	A75585	transcription regu
688	6	0.6	223	2	AH0409	761	6	0.6	239	2	B95374	hypothetical prote
689	6	0.6	224	2	E83859	762	6	0.6	240	1	Z12M92	19k zein precursor
690	6	0.6	224	2	H71203	763	6	0.6	240	2	B99657	probable oxidoredu
691	6	0.6	224	2	G75218	764	6	0.6	240	2	F85517	probable oxidoredu
692	6	0.6	224	2	A64418	765	6	0.6	240	2	D95420	probable oxidoredu
693	6	0.6	224	2	T51638	766	6	0.6	240	2	A22831	19k zein precursor
694	6	0.6	225	2	G64702	767	6	0.6	240	2	H71435	hypothetical prote
695	6	0.6	225	2	F71817	768	6	0.6	241	2	C90532	ribonuclease III (
696	6	0.6	226	2	D87061	769	6	0.6	241	2	F71478	probable pseudoini
697	6	0.6	226	2	C82124	770	6	0.6	241	2	E81741	ribosomal large ch
698	6	0.6	226	2	D97786	771	6	0.6	241	2	E64386	conserved hypotnet
699	6	0.6	227	2	AB1076	772	6	0.6	241	2	A75065	hypothetical prote
700	6	0.6	227	2	A71224	773	6	0.6	241	2	D71167	hypothetical prote
701	6	0.6	227	2	H75191	774	6	0.6	241	2	A86651	glycosyltransferas
702	6	0.6	227	2	E84403	775	6	0.6	242	2	G82206	hypothetical prote
703	6	0.6	227	2	D83543	776	6	0.6	242	2	E88250	protein T21B10.5 (
704	6	0.6	227	2	H69845	777	6	0.6	242	2	F84315	cohalat transport A
705	6	0.6	228	2	S76876	778	6	0.6	243	2	S46732	hypothetical prote
706	6	0.6	228	2	S34646	779	6	0.6	243	2	S56405	hypothetical 26.6k
707	6	0.6	228	2	T32838	780	6	0.6	243	2	D91273	hypothetical prote
708	6	0.6	228	2	AD3326	781	6	0.6	243	2	D86114	hypothetical prote
709	6	0.6	229	2	H64367	782	6	0.6	243	2	A11049	probable tRNA/rRNA
710	6	0.6	229	2	T06204	783	6	0.6	243	2	T12485	hypothetical prote
711	6	0.6	229	2	F70568	784	6	0.6	243	2	B95885	hypothetical prote
712	6	0.6	230	2	D58892	785	6	0.6	244	2	S50685	hypothetical prote
713	6	0.6	230	2	A6387	786	6	0.6	244	2	F95262	hypothetical prote
714	6	0.6	230	2	S34645	787	6	0.6	244	2	AB3524	transcription regu
715	6	0.6	230	2	AF3113	788	6	0.6	245	2	S57550	hypothetical prote
716	6	0.6	231	2	T03765	789	6	0.6	245	2	F71887	hypothetical prote
717	6	0.6	231	2	A38083	790	6	0.6	245	2	A72275	hypothetical prote
718	6	0.6	231	2	AD0325	791	6	0.6	245	2	H82202	hypothetical prote
719	6	0.6	232	2	F64363	792	6	0.6	246	2	G82819	acetoacetyl-CoA re
720	6	0.6	232	2	D85654	793	6	0.6	246	2	F70062	hypothetical prote
721	6	0.6	232	2	A97748	794	6	0.6	246	2	E64627	hypothetical prote
722	6	0.6	232	2	B90230	795	6	0.6	246	2	E69230	hypothetical prote
723	6	0.6	232	2	F97801	796	6	0.6	246	2	T12585	DC3 promoter-bindi
724	6	0.6	233	2	E72738	797	6	0.6	246	2	B90224	hypothetical prote
725	6	0.6	233	2	G72527	798	6	0.6	246	2	A43579	Vmp7 protein homol
726	6	0.6	234	2	S28482	799	6	0.6	247	2	A99794	urase-associated
727	6	0.6	234	2	T30427	800	6	0.6	247	2	S67685	hypothetical prote
728	6	0.6	234	2	T21543	801	6	0.6	247	2	H82348	hypothetical prote
729	6	0.6	234	2	G98173	802	6	0.6	247	2	H86844	glutamine ABC tran
730	6	0.6	235	2	T08467	803	6	0.6	247	2	T14818	leucine-rich repaa
731	6	0.6	235	2	T44679	804	6	0.6	248	1	S52234	3',5'-cyclic-nucle
732	6	0.6	235	2	A98286	805	6	0.6	248	2	AC3010	gramicidin S biosy
733	6	0.6	235	2	H70848	806	6	0.6	248	2	G83478	probable permease
734	6	0.6	235	2	T20337	807	6	0.6	248	2	A98143	response regulator
735	6	0.6	235	2	E86127	808	6	0.6	248	2	AB3145	two component resp
736	6	0.6	235	2	AB1255	809	6	0.6	248	2	C97439	sucrose-phosphatas
737	6	0.6	236	2	T22003	810	6	0.6	248	2	AF2657	hydrolyase (limpor
738	6	0.6	236	2	I46860	811	6	0.6	249	2	T21169	hypothetical prote
739	6	0.6	236	2	S17932	812	6	0.6	249	2	T52142	RING finger protei
740	6	0.6	236	2	B70453	813	6	0.6	249	2	AB3799	hypothetical prote
741	6	0.6	236	2	T43237	814	6	0.6	249	2	D85693	unknown protein en
742	6	0.6	236	2	C69060	815	6	0.6	250	2	T25582	hypothetical prote
743	6	0.6	236	2	C69396	816	6	0.6	250	2	F70153	conserved hypotnet
744	6	0.6	236	2	AE2268	817	6	0.6	250	2	B91020	hypothetical prote
745	6	0.6	237	2	T47012	818	6	0.6	250	2	D85864	hypothetical prote
746	6	0.6	237	2	AD2567	819	6	0.6	250	2	AC2026	hypothetical prote
747	6	0.6	237	2	S65313	820	6	0.6	250	2	F86761	nucleotidase (limp
748	6	0.6	237	2	G70066	821	6	0.6	251	2	C83053	triasephosphatase
749	6	0.6	237	2	H81083	822	6	0.6	251	2	AB3390	is arginase/ornithine
750	6	0.6	237	2	D87373	823	6	0.6	251	2	T11549	AMP binding protei
751	6	0.6	237	2	T36804	824	6	0.6	251	2	T14456	MAOS box protein h
752	6	0.6	237	2	T24407	825	6	0.6	251	2	S73162	hypothetical prote
753	6	0.6	237	2	AE0237	826	6	0.6	251	2	G97579	ABC transporter AT
754	6	0.6	238	2	S41313	827	6	0.6	251	2	AE2800	hypothetical prote
755	6	0.6	238	2	H81859	828	6	0.6	251	2	H87692	conserved hypotnet
756	6	0.6	238	2	AE2027	829	6	0.6	251	2	G75345	serine/threonine-p
757	6	0.6	238	2	C81145	830	6	0.6	251	2	A10100	probable carbonic
758	6	0.6	239	2	T45805	831	6	0.6	251	2	AH2181	hypothetical prote
759	6	0.6	239	2	G81369	832	6	0.6	252	2	F83098	probable short-cha

833	6	0.6	252	2	146859	MHC RLA - rabbit (906	6	0.6	266	2	B39260	MHC class II histo
834	6	0.6	252	2	F39534	floral homeotic pr	907	6	0.6	266	2	E64955	flly protein precu
835	6	0.6	252	2	A95071	amino acid ABC tra	908	6	0.6	266	2	C90961	probable periplasm
836	6	0.6	252	2	A34877	C4b-binding protel	909	6	0.6	266	2	C85809	probable periplasm
837	6	0.6	252	2	JC2399	pms4 homolog misma	910	6	0.6	266	2	A46552	chlorophyll a/b-bl
838	6	0.6	252	2	AH1482	hypothetical prote	911	6	0.6	266	2	H69189	hypothetical prote
839	6	0.6	253	2	TS0942	Ditb protein (limp	912	6	0.6	266	2	E71612	ribosomal protein
840	6	0.6	253	2	JC5014	2,4-dichlorophenol	913	6	0.6	267	2	C64995	hypothetical prote
841	6	0.6	253	2	SA9055	thiesterase - Str	914	6	0.6	267	2	AH0792	probable 2,4-dihyd
842	6	0.6	253	2	S28541	RNA N-glycosidase	915	6	0.6	267	2	I40327	bat protein - Bord
843	6	0.6	253	2	S28539	RNA N-glycosidase	916	6	0.6	268	2	F86680	prophage pil prote
844	6	0.6	253	2	S28542	RNA N-glycosidase	917	6	0.6	268	2	SA5091	hypothetical prote
845	6	0.6	253	2	S29931	RNA N-glycosidase	918	6	0.6	268	2	E55607	2-oxo-hepta-3-ene-
846	6	0.6	253	2	C36789	hypothetical prote	919	6	0.6	268	2	A64361	hypothetical prote
847	6	0.6	253	2	A90558	hypothetical prote	920	6	0.6	269	2	DS0171	chlorophyll a/b-bl
848	6	0.6	254	2	C98274	nikf2 protein (Auj	921	6	0.6	269	2	D90141	hypothetical prote
849	6	0.6	254	2	F90107	60S ribosomal prot	922	6	0.6	269	2	I39961	hypothetical prote
850	6	0.6	254	2	A64437	hypothetical prote	923	6	0.6	270	2	D70171	ABC transporter, A
851	6	0.6	255	2	T12074	urate oxidase (EC	924	6	0.6	270	2	AH2455	hypothetical prote
852	6	0.6	255	2	AB3500	phosphate regulon	925	6	0.6	271	2	G89809	hypothetical prote
853	6	0.6	255	2	T18206	mxd protein - Bac	926	6	0.6	271	2	D83059	phosphatidylserine
854	6	0.6	255	2	T03299	expansin 3 - rice	927	6	0.6	271	2	AC3527	aldenhyde dehydrog
855	6	0.6	255	2	C75527	conserved hypotet	928	6	0.6	272	2	S77576	oligopeptide trans
856	6	0.6	255	2	AE3551	tetratricopeptide	929	6	0.6	272	2	D86482	protein F5j5.3 (lm
857	6	0.6	256	2	D97086	formate/nitrite fa	930	6	0.6	272	2	T25044	hypothetical prote
858	6	0.6	256	2	AB0348	hypothetical prote	931	6	0.6	272	2	AC1916	hypothetical prote
859	6	0.6	257	2	A10935	acetylglutamate kl	932	6	0.6	273	1	H69744	conserved hypotet
860	6	0.6	257	2	T64159	hypothetical prote	933	6	0.6	273	2	E72231	ATP synthase Fl, s
861	6	0.6	257	2	B90194	phosphate ABC tran	934	6	0.6	273	2	T48810	pall related prote
862	6	0.6	257	2	S33314	beta-etherase ligf	935	6	0.6	273	2	T12793	hypothetical prote
863	6	0.6	257	2	D70341	hypothetical prote	936	6	0.6	273	2	B86788	Lysr family transc
864	6	0.6	258	1	VHVUM1	nucleoprotein N -	937	6	0.6	273	2	T19174	hypothetical prote
865	6	0.6	258	1	VHVUMC	nucleoprotein N -	938	6	0.6	273	2	AB0848	Iron transport pro
866	6	0.6	258	1	JQ1876	nucleoprotein N -	939	6	0.6	274	2	AB3583	hypothetical prote
867	6	0.6	258	1	S39747	ywfN protein - Bac	940	6	0.6	274	2	H88690	protein F4IH10.8 l
868	6	0.6	258	1	T36303	probable indolegly	941	6	0.6	274	2	F83601	hypothetical prote
869	6	0.6	258	2	S17105	nucleoprotein N -	942	6	0.6	274	2	B70901	hypothetical prote
870	6	0.6	258	2	SA7242	gene N protein - t	943	6	0.6	275	2	A87487	phosphatidate cytl
871	6	0.6	258	2	A81950	hypothetical prote	944	6	0.6	275	2	AD3383	hypothetical prote
872	6	0.6	259	2	T44461	transcription init	945	6	0.6	275	2	T43119	hypothetical prote
873	6	0.6	259	2	S32432	toxln cytb - Bacil	946	6	0.6	275	2	E93231	pur operon repress
874	6	0.6	260	1	CRMS2	carbonate dehydrat	947	6	0.6	275	2	G98095	activator of purin
875	6	0.6	260	2	H81297	Fdhb protein Cj150	948	6	0.6	276	2	T09131	chitinase (EC 3.2.
876	6	0.6	261	2	D71508	probable glutamine	949	6	0.6	276	2	G69023	nucleotide-binding
877	6	0.6	261	2	T16650	hypothetical prote	950	6	0.6	276	2	AB4361	hypothetical prote
878	6	0.6	262	1	SA8029	enoyl-lacyl-carrie	951	6	0.6	276	2	C88650	protein C09G12.3 l
879	6	0.6	262	1	B43729	enoyl-lacyl-carrie	952	6	0.6	276	2	D83786	glycerol uptake fa
880	6	0.6	262	2	E90861	enoyl-lacyl-carrie	953	6	0.6	277	2	S34146	porin porI, plasti
881	6	0.6	262	2	H85757	enoyl-lacyl-carrie	954	6	0.6	277	2	S75973	hypothetical prote
882	6	0.6	262	2	AD0656	enoyl-lacyl-carrie	955	6	0.6	277	2	C90077	conserved hypotet
883	6	0.6	262	2	C97651	psb protein (limp	956	6	0.6	278	2	C43670	integral membrane
884	6	0.6	262	2	C75051	hydrolyase related	957	6	0.6	278	2	S77601	hypothetical prote
885	6	0.6	262	2	G98343	hypothetical prote	958	6	0.6	278	2	T08252	probable transposa
886	6	0.6	262	2	AH2938	hypothetical prote	959	6	0.6	279	1	S0MRTV	thermistase (EC 3.4
887	6	0.6	262	2	A65220	phosphonates trans	960	6	0.6	279	2	G95195	sugar ABC transpor
888	6	0.6	262	2	H91264	ATP-binding compo	961	6	0.6	279	2	D98062	hypothetical prote
889	6	0.6	262	2	B97070	precorrin-6x reduc	962	6	0.6	279	2	T37148	probable DNA-bindl
890	6	0.6	262	2	E86105	ATP-binding compo	963	6	0.6	280	2	AB2875	exopolysaccharide
891	6	0.6	263	2	C81667	amino acid ABC tra	964	6	0.6	280	2	E71288	probable hydrogena
892	6	0.6	263	2	E97774	hypothetical prote	965	6	0.6	280	2	C91175	hypothetical prote
893	6	0.6	263	2	T16399	hypothetical prote	966	6	0.6	280	2	C86021	hypothetical prote
894	6	0.6	263	2	S73719	MG313 homolog H08	967	6	0.6	280	2	A48488	saccharide biosynt
895	6	0.6	263	2	AE0140	molybdenum transp	968	6	0.6	281	2	T09653	delta-1-pyrrolone-
896	6	0.6	264	2	T31271	2-hydroxypent-2,4	969	6	0.6	281	2	SA1196	transcription regu
897	6	0.6	264	2	A71367	probable holocytoe	970	6	0.6	281	2	S34545	hypothetical prote
898	6	0.6	264	2	F72577	hypothetical prote	971	6	0.6	281	2	E84637	hypothetical prote
899	6	0.6	265	2	T13116	protein gp30 - pha	972	6	0.6	281	2	S63225	ribosomal protein
900	6	0.6	265	2	H81155	conserved hypotet	973	6	0.6	281	2	F82670	general secretory
901	6	0.6	265	2	D90320	hypothetical prote	974	6	0.6	281	2	AE3523	icc protein (limpor
902	6	0.6	265	2	D83287	hypothetical prote	975	6	0.6	282	2	C90016	hypothetical prote
903	6	0.6	265	2	AS9141	silaflin slilp pre	976	6	0.6	282	2	E82352	conserved hypotet
904	6	0.6	266	1	F71701	hypothetical prote	977	6	0.6	282	2	F69880	conserved hypotet
905	6	0.6	266	2	A39260	MHC class II histo	978	6	0.6	282	2	S47968	casein kinase II (

```

979 6 0.6 282 2 H83651 hypothetical prote
980 6 0.6 282 2 B71601 probable integral
981 6 0.6 282 2 T50833 nematode resistanc
982 6 0.6 283 2 S05205 rRNA N-glycosidase
983 6 0.6 283 2 G70376 conserved hypothet
984 6 0.6 283 2 T27423 hypothetical prote
985 6 0.6 283 2 F91285 hypothetical prote
986 6 0.6 283 2 E49348 succinoglycan bios
987 6 0.6 284 2 AH3570 nickel transport s
988 6 0.6 284 2 B64694 conserved hypothet
989 6 0.6 284 2 T39813 hypothetical prote
990 6 0.6 284 2 E95008 conserved hypothet
991 6 0.6 284 2 B97880 conserved hypothet
992 6 0.6 284 2 A70976 hypothetical prote
993 6 0.6 284 2 E82868 hypothetical prote
994 6 0.6 284 2 AD0987 probable phosphosu
995 6 0.6 285 2 C97279 thioedoxin reduct
996 6 0.6 285 2 I50995 connexin 32.2 - At
997 6 0.6 285 2 G83233 conserved hypothet
998 6 0.6 285 2 T21611 hypothetical prote
999 6 0.6 286 2 AE1200 3-hydroxyisobutyra
1000 6 0.6 286 2 D95401 probable oxidoredu

```

ALIGNMENTS

```

RESULT 1
T7880
hypothetical protein ZK520.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T7880
R:Steward, C.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z20434
A:Accession: T7880
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-519 <MW>
A:Cross-references: EMBL:Z92822; PIDN:CA807299.1; GSPDB:GN00021; CESP:ZK520.1
A:Experimental source: clone ZK520
C:Genetics:
A:Gene: CESP:ZK520.1
A:Map position: 3
A:Introns: 12/1; 69/3; 106/2; 178/3; 240/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK520.1

Query Match 0.8%; Score 9; DB 2; Length 519;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1009 SAIOFLVMS 1017
DB 293 SAIOFLVMS 301

RESULT 2
T17821
hypothetical protein A322L - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T17821
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T17821
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-176 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AC96690.1
A:Experimental source: specific host Chlorella strain NC64A

```

```

C:Genetics:
A:Note: A322L
C:Superfamily: Chlorella virus PBCV-1 hypothetical protein A322L

```

```

Query Match 0.7%; Score 8; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 628 KDTIOGAK 635
DB 103 KDTIOGAK 110

```

```

RESULT 3
A84283
hypothetical protein Vng1276c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84283
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: A84283
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-241 <STO>
A:Cross-references: GB:AE004437; NID:g10580799; PIDN:AA619629.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1276C

```

```

Query Match 0.7%; Score 8; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 788 AALEMRD 795
DB 27 AALEMRD 34

```

```

RESULT 4
AE3648
flagellar motor switch protein flm [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AE3648
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD3252; PMID:1175688
A:Accession: AE3648
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-316 <KUR>
A:Cross-references: GB:AE008918; PIDN:AA154352.1; PID:g17985334; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11110
A:Map position: II

```

```

Query Match 0.7%; Score 8; DB 2; Length 316;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 760 NLAGHIA 767
DB 760 NLAGHIA 767

```


Db 135 NLAGH14 142

RESULT 5

KIBSGM

phosphoglycerate kinase (EC 2.7.2.3) - *Bacillus megaterium*C:Species: *Bacillus megaterium*

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 21-Jan-2000

C:Accession: S13125; J01954

R:Schlaepfer, B.S.; Brniant, G.; Zuber, H.

Nucleic Acids Res. 18, 6423, 1990

A:Title: Nucleotide sequence of the phosphoglycerate kinase gene from *Bacillus megaterium*

A:Reference number: S13125; M01D:91057129

A:Accession: S13125

A:Molecule type: DNA

A:Residues: 1-394 <SCH>

A:Cross-references: EMBL:X54519; NID:g39642; PIDN:CAA38375.1; PID:g39643

A:Experimental source: DSM 319

R:Schlaepfer, B.S.; Zuber, H.

Gene 122, 53-62, 1992

A:Title: Cloning and sequencing of the genes encoding glyceraldehyde-3-phosphate dehydrogenase and corresponding sequences from thermophilic *Bacillus stearothermophilus*.

A:Reference number: J01952; M01D:93083995

A:Accession: J01954

A:Molecule type: DNA

A:Residues: 1-394 <SC2>

A:Cross-references: GB:M87647; NID:g143315; PIDN:AA73203.1; PID:g143318

A:Experimental source: strain DSM319

C:Genetics:

A:Gene: pgk

C:Superfamily: phosphoglycerate kinase

C:Keywords: ATP; gluconeogenesis; glycolysis; phosphotransferase

F:201,323/Binding site: ATP (Lys, Glu) #status predicted

Query Match

Best Local Similarity 0.7%; Score 8; DB 1; Length 394;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 632 QGAKV14 639

Db 50 QGAKV14 57

RESULT 6

AC0575

hypothetical membrane protein p43 STY0637 [imported] - *Salmonella enterica* subsp. *enterica*C:Species: *Salmonella enterica* subsp. *enterica* serovar TyphimuriumA:Note: this species has also been called *Salmonella typhimurium*

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001

C:Accession: AC0575

R:Perkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, C.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar

A:Reference number: AB0502; PMID:11677608

A:Accession: AC0575

A:Molecule type: DNA

A:Residues: 1-414 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05069.1; PID:g16501844; GSPDB:GN00176

C:Genetics:

A:Gene: STY0637

C:Superfamily: Streptomyces lividans chloramphenicol resistance protein

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 414;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 VGSFLAVG 155

Db 296 VGSFLAVG 303

RESULT 7

E87233

probable secreted protein [imported] - *Mycobacterium leprae*C:Species: *Mycobacterium leprae*

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: E87233

R:Coile, S.T.; Elismeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; R.; Davies, R.M.; Devlin, K.; Dutfoy, S.; Fellwell, T.; Fraser, A.; Hamlin, N.; Holto

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; M01D:21128732; PMID:11234002

A:Accession: E87233

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-519 <SMO>

A:Cross-references: GB:AL450380; NID:g13093825; PIDN:CAC32123.1; GSPDB:GN00147

C:Genetics:

A:Gene: mceIC

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 519;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 561 FPDNGTR 568

Db 346 FPDNGTR 353

RESULT 8

T29967

hypothetical protein F36A4.15 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000

C:Accession: T29967

R:Miller, N.; Bradshaw, H.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F36A4.

A:Reference number: Z20713

A:Accession: T29967

A:Status: preliminary; translated from GB/EMBL/DDBT

A:Molecule type: DNA

A:Residues: 1-642 <ML>

A:Cross-references: EMBL:U53333; PIDN:AA96166.1; GSPDB:GN00022; CESP:F36A4.15

A:Experimental source: strain Bristol N2; clone F36A4

C:Genetics:

A:Gene: CESP:F36A4.15

A:Map position: 4

A:Insertions: 14/3; 116/2; 146/3; 181/2; 388/3; 466/2

C:Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 642;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1057 FEGERVYL 1064

Db 282 FEGERVYL 289

RESULT 9

T34618

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain nuog - *Streptomyces coelicolor*C:Species: *Streptomyces coelicolor*

C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 04-Mar-2000

C:Accession: T34618

R.Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, June 1999
A:Reference number: 221549
A:Accession: T34618
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-843 <SAU>
A:Cross-references: EMBL:AL078618; PIDN:CAB44525.1; GSPDB:GN00070; SCOEDB:SC10A7.15C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC10A7.15C
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 11
C:Keywords: NAD; oxidoreductase

Query Match 0.7%; Score 8; DB 2; Length 843;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 KDGVDLAV 111
|||||||
DB 775 KDGVDLAV 782

RESULT 10
T47659
splicosomal-like protein - Arabidopsis thaliana
N:Alternat names: protein T26112.100; protein T26112.80
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
R:Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24471
A:Accession: T47659
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1214 <MON>
A:Cross-references: EMBL:AL132954
A:Experimental source: cultivar Columbia; BAC clone T26112
A:Accession: T47661
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1214 <MO2>
A:Cross-references: EMBL:AL132954
A:Experimental source: cultivar Columbia; BAC clone T26112
C:Genetics:
A:Map position: 3
A:Introns: 1005/3; 1169/3
A:Note: T26112.80; T26112.100

Query Match 0.7%; Score 8; DB 2; Length 1214;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 152 LAVGVK 159
|||||||
DB 902 LAVGVK 909

RESULT 11
T32625
hypothetical protein F15E6.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T32625
R:Miller, N.; Stellyes, L.; Bradshaw, H.; Keppler, D.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F15E6.
A:Reference number: Z21202
A:Accession: T32625
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-1238 <ML>
A:Cross-references: EMBL:AF038614; PIDN:AB92058.1; GSPDB:GN00022; CESP:F15E6.6
A:Experimental source: strain Bristol NZ; clone F15E6
C:Genetics:
A:Gene: CESP:F15E6.6
A:Map position: 4
A:Introns: 34/1; 109/3; 225/3; 271/1; 312/1; 345/3; 412/3; 517/3; 551/1; 616/3; 709/3
C:Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology
C:Keywords: 2Fe-2S; metalloprotein
F:43/48,51,73/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 0.7%; Score 8; DB 2; Length 1238;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1057 FEGERKYL 1064
|||||||
DB 878 FEGERKYL 885

RESULT 12
G98275
hypothetical protein AGR_L_2329 [imported] - Agrobacterium tumefaciens (strain C58, C
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: G98275
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Gold
A.; Liu, F.; Moliam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Matkeltz,
Science 294, 2353-2358, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: G98275
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1520 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89729.1; PID:q15159645; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_2329
A:Map position: linear chromosome

Query Match 0.7%; Score 8; DB 2; Length 1520;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 633 GAKVILAG 640
|||||||
DB 1175 GAKVILAG 1182

RESULT 13
AF3008
polyketide synthase Atu3672 [imported] - Agrobacterium tumefaciens (strain C58, Dupon
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AF3008
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuyavlin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF3008
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1520 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA144484.1; PID:q17742089; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:

A:Gene: Atu3672
A:Map position: linear chromosome

Query Match 0.7%; Score 8; DB 2; Length 1520;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 633 GAKVILAG 640
|||||
DB 1175 GAKVILAG 1182

RESULT 14

A96026
probable transposase of insertion sequence ISKm2011-2, orfB N-terminus protein [imported]
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: A96026
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmaster, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: A96026
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-39 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49874.1; PID:g15141362; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Publer, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowle, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebulit, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Trm2011-2b-1; Smb21702
A:Genome: plasmid

Query Match 0.7%; Score 7; DB 2; Length 39;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 568 RLVFIDE 574
|||||
DB 32 RLVFIDE 38

RESULT 15

S14968
hypothetical NADH dehydrogenase chain 4 homolog - Synechocystis sp. (strain PCC 6803)
N:Alternate names: ndhd protein-like protein
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 15-Oct-1998 #sequence_revision 15-Oct-1998 #text_change 02-Jun-2000
C:Accession: S14968
R:Anderson, S.L.; McIntosh, L.
Plant Mol. Biol. 16, 487-499, 1991
A:Title: Partial conservation of the 5' ndhE-psac-ndhD 3' gene arrangement of chloroplast
oplasts.
A:Reference number: S14966; MUID:91329685
A:Accession: S14968
A:Molecule type: DNA
A:Residues: 1-91 <AND>
A:Cross-references: EMBL:X53842; NID:g47594; PIDN:CA37837.1; PID:g581748
A:Experimental source: PCC 6803
A:Note: the authors translated the initiation codon GTG as Val
C:Comment: This is the hypothetical translation of a sequence that was not reported as a
co chloroplast (see PIR:DEMTN4).

C:Genetics:
A:Start codon: GTG

Query Match 0.7%; Score 7; DB 4; Length 91;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 230 KMDDRTS 236
|||||
DB 62 KMDDRTS 68

Search completed: May 21, 2002, 14:55:07
Job time: 186 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 14:52:21 : Search time 15.19 Seconds
(without alignments)
2737.643 Million cell updates/sec

Title: US-09-729-653-2_COPY_1_1074

Perfect score: 1074

Sequence: 1 HSLIGRCSRLGIDGNVAVC.....LYFEGEKRYLOAGKFFLLCG 1074

Scoring table: OLIGO

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 1000 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	8	0.7	394 1	PGK_BACME
2	8	0.7	843 1	NUOG_STRCO
3	7	0.7	15 1	LCK_DROME
4	7	0.7	120 1	YACL_ECOLI
5	7	0.7	131 1	RT17_YEAST
6	7	0.7	144 1	RL15_ECOLI
7	7	0.7	144 1	RL15_HAEIN
8	7	0.7	157 1	FMO_MORBO
9	7	0.7	159 1	FMI_MORBO
10	7	0.7	159 1	MLE_TODPA
11	7	0.7	173 1	POPS_YEAST
12	7	0.7	182 1	UCRT_RHORI
13	7	0.7	190 1	YEIP_ECOLI
14	7	0.7	196 1	SODE_TETPY
15	7	0.7	217 1	DFRA_MEDSA
16	7	0.7	219 1	R1B7_ARCFU
17	7	0.7	225 1	CSMJ_CHLTE
18	7	0.7	249 1	BLAB_BACFR
19	7	0.7	262 1	MURI_BUCAI
20	7	0.7	267 1	YG95_HAEIN
21	7	0.7	287 1	V033_FOWPV
22	7	0.7	295 1	RFBA_XANCP
23	7	0.7	302 1	MEUR_ECOLI
24	7	0.7	331 1	LEUI_THETH
25	7	0.7	332 1	TWFI_YEAST
26	7	0.7	336 1	G3P_ASPNG
27	7	0.7	337 1	G3P_CRYPA
28	7	0.7	338 1	G3P_NEUCR
29	7	0.7	351 1	MTD5_DACSA
30	7	0.7	359 1	VALI_TYLCM
31	7	0.7	359 1	VALI_TYLCU
32	7	0.7	364 1	VMAT_NDVA
33	7	0.7	364 1	VMAT_NDVB

34	7	0.7	370 1	VE2_HPV39	P24630 human papil
35	7	0.7	386 1	RIR2_BRARE	P79733 brachydanio
36	7	0.7	394 1	PGK_BACST	P18912 bacillus st
37	7	0.7	396 1	DXR_DEIRA	O9RU84 delnoccocus
38	7	0.7	401 1	PGKT_WHEAT	P12783 triticum ae
39	7	0.7	402 1	DFP_STNY3	P73881 synechocyst
40	7	0.7	443 1	SGN2_HUMAN	Q15647 homo sapien
41	7	0.7	448 1	TX19_HUMAN	O60806 homo sapien
42	7	0.7	450 1	SR54_MYCPN	P75054 mycoplasma
43	7	0.7	485 1	GLME_CLOTT	O05509 clostridium
44	7	0.7	535 1	D2_DICDI	P18142 dicystostell
45	7	0.7	536 1	PGCK_RHIME	P43085 rhizobium m
46	7	0.7	547 1	GUNL_BUTFI	P20847 butyryllybri
47	7	0.7	566 1	ROCB_BACSU	P39635 bacillus su
48	7	0.7	574 1	DPO1_AOUAE	O67779 aquifex aeo
49	7	0.7	576 1	PEXS_PICPA	P33392 pichia past
50	7	0.7	589 1	RESE_BACSU	P35164 bacillus su
51	7	0.7	612 1	GIDA_MYCGE	P47619 mycoplasma
52	7	0.7	629 1	HTPQ_RHIME	P58477 rhizobium m
53	7	0.7	632 1	DXS_CHLMU	O9PK62 chlamydia m
54	7	0.7	640 1	DXS_CHLTR	O84335 chlamydia t
55	7	0.7	654 1	PGKT_THEMA	P36304 thermotoga
56	7	0.7	695 1	EBG1_TREPA	O83748 treponema p
57	7	0.7	702 1	SPB1_BRAJU	O82475 brassica ju
58	7	0.7	745 1	AXN_DROME	O9V407 drosophila
59	7	0.7	816 1	YM35_YEAST	O03823 saccharomyc
60	7	0.7	889 1	COPE_YEAST	P41811 saccharomyc
61	7	0.7	906 1	HEIX_MYCTU	Q10701 mycobacteri
62	7	0.7	957 1	NIRB_KLEPN	O06458 klebsiella
63	7	0.7	988 1	DPOL_FOWPV	P21402 fowlpox vir
64	7	0.7	993 1	YIS2_YEAST	P40562 saccharomyc
65	7	0.7	1188 1	PPSA_METJA	O57662 methanococc
66	7	0.7	1296 1	BXG_CLOBO	O60393 clostridium
67	7	0.7	1403 1	PRO_DROME	P29617 drosophila
68	7	0.7	1612 1	TP2B_CRITO	O64399 citreulius
69	7	0.7	1612 1	TP2B_MOUSE	O64511 homo musculu
70	7	0.7	1626 1	TP2B_HUMAN	O02880 homo sapien
71	7	0.7	1643 1	OMPA_RICPR	O53020 r outer mem
72	7	0.7	2090 1	N214_HUMAN	P35658 homo sapien
73	7	0.7	3305 1	APLP_MANSE	O25490 manduca sex
74	6	0.6	15 1	LEC3_AXIPO	P28588 axinella po
75	6	0.6	54 1	BAF_BORBR	O08005 boridietella
76	6	0.6	56 1	YPMA_BACSU	P54395 bacillus su
77	6	0.6	63 1	MT1_COLLI	P15786 columbia liv
78	6	0.6	66 1	PSAE_FREDI	P236075 fremyella d
79	6	0.6	67 1	CSRA_MECIU	O30875 micrococcu
80	6	0.6	69 1	PSAE_STNP2	P31369 synechococc
81	6	0.6	70 1	PSAE_CYAPA	P48114 cyanophora
82	6	0.6	85 1	DYLI_SCHPO	O9U005 schizosacch
83	6	0.6	87 1	F115_MOUSE	O61075 mus musculu
84	6	0.6	87 1	YD63_ARCFU	O28908 archaeoglob
85	6	0.6	88 1	ATP9_DICDI	O37315 dicystostell
86	6	0.6	88 1	PMRD_ECOLI	P37390 escherichia
87	6	0.6	99 1	Y142_UREPA	O9P102 ureaplasma
88	6	0.6	101 1	VE4_HPV41	P27553 human papil
89	6	0.6	105 1	GLNI_METBA	P54508 methanosarc
90	6	0.6	107 1	FKB1_BOVIN	P20071 bos tauris
91	6	0.6	107 1	FKB1_HUMAN	P26883 mus musculu
92	6	0.6	107 1	FKB1_MOUSE	O62658 rattus norv
93	6	0.6	107 1	FKB1_XENLA	O42123 xenopus lae
94	6	0.6	107 1	FKB1_XENLA	O42123 xenopus lae
95	6	0.6	107 1	GUAN_CAVPO	P70664 cavia porce
96	6	0.6	108 1	KV3A_HUMAN	P01619 homo sapien
97	6	0.6	109 1	CLP1_HELAM	O02443 helicoverpa
98	6	0.6	109 1	GUAN_PIG	P79697 sus scrofa
99	6	0.6	111 1	YW12_STRCO	P23159 streptomyc
100	6	0.6	115 1	GUAN_HUMAN	O02747 homo sapien
101	6	0.6	116 1	SMS_CHICK	P33094 gallus gall
102	6	0.6	116 1	WN5B_ALOUV	P28104 alopas vul
103	6	0.6	116 1	WN5B_EUMSK	P28118 eumeces ski
104	6	0.6	118 1	POL_HVICA	P05360 human immun
105	6	0.6	119 1	FOLX_ECOLI	P80449 escherichia
106	6	0.6	121 1	RS10_MYCGA	O52331 mycoplasma

PK	PK_NAME	STANDARD	PRT	394 AA
983	6	0.6	727	1
984	6	0.6	727	1
985	6	0.6	729	1
986	6	0.6	729	1
987	6	0.6	729	1
988	6	0.6	729	1
989	6	0.6	734	1
990	6	0.6	736	1
991	6	0.6	737	1
992	6	0.6	737	1
993	6	0.6	738	1
994	6	0.6	739	1
995	6	0.6	739	1
996	6	0.6	739	1
997	6	0.6	740	1
998	6	0.6	745	1
999	6	0.6	746	1
1000	6	0.6	748	1

ALIGNMENTS

RESULT 1

PK_BACME STANDARD; PRT: 394 AA.

ID PK_BACME STANDARD; PRT: 394 AA.

AC P24269;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Phosphoglycerate kinase (EC 2.7.2.3).

GN PGK.

OS Bacillus megaterium.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1404;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-DSM 319;

RA MEDLINE-91057129; PubMed-213031;

RA Schlaepfer B.S., Brantant C., Brantant G., Zuber H.;

RT "Nucleotide sequence of the phosphoglycerate kinase gene from

RT Bacillus megaterium";

RL Nucleic Acids Res. 18:6423-6423(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-DSM 319;

RX MEDLINE-93083995; PubMed-1452037;

RA Schlaepfer B.S., Zuber H.;

RT "Cloning and sequencing of the genes encoding

RT glyceralddehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and

RT triosephosphate isomerase (gap operon) from mesophilic Bacillus

RT megaterium: comparison with corresponding sequences from thermophilic

RT Bacillus stearothermophilus.";

RL Gene 122:53-62(1992)

RT CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-

CC -1- PARTWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.

CC -1- SUBUNIT: MONOMER.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

DR EMBL: X54519; CAA38375.1; -

DR EMBL: M87647; AAA73203.1; -

DR EMBL: M87648; AAA73206.1; -

DR PIR: S13125; KIBSGM.

DR PIR: J01954; J01954.

DR HSSP: P18912; 1PHD.

DR InterPro: IPR001576; PGK.

DR Pfam: PF00162; PGK.1.

DR PRINTS: PR00477; PHGLCKINASE.

DR PROSITE: PS00111; GLYCERATE_KINASE.1.

DR Transferase: Kinase; Glycolysis.

SQ SEQUENCE 394 AA; 42457 MW; 7A2E6B978FA7008B CRC64;

Query Match 0.7%; Score 8; DB 1; Length 394;

Best Local Similarity 100.0%; Pred. No. 5.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 632 OGAKVILA 639

DB 50 OGAKVILA 57

RESULT 2

NUOG_STRCO STANDARD; PRT: 843 AA.

ID NUOG_STRCO STANDARD; PRT: 843 AA.

AC 09XAR0;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE NADH dehydrogenase I chain G (EC 1.6.5.3) (NADH-ubiquinone

DE oxidoreductase chain G).

GN NUOG OR SC10A7.15C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Saunders D.C., Harris D., Bentley S.D., Parkhill J., Barrell B.G.,

RA Rajandream M.A.;

RL Submitted (JUN-1999) to the EMBL/GenBank/DBA databases.

CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

CC -1- COFACTOR: BINDS 1 2FE-2S CLUSTER AND 1 4FE-4S CLUSTER (potential).

CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 75 kDa SUBUNIT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

DR EMBL: AL07618; CAB44525.1; -

DR InterPro: IPR000283; ComplexI_75k.

DR InterPro: IPR001041; Ferredoxin.

DR InterPro: IPR001467; Molybdopterin.

DR Pfam: PF00384; molybdopterin; 2.

DR PROSITE: PS00641; COMPLEX1_75K.1; 1.

DR PROSITE: PS00642; COMPLEX1_75K.2; 1.

DR PROSITE: PS00643; COMPLEX1_75K.3; 1.

KW Oxidoreductase; NAD; Ubiquinone; Iron-sulfur; 4Fe-4S.

FT METAL 54

FT METAL 54

FT METAL 65

FT METAL 68

FT METAL 120

FT METAL 120

FT METAL 123

FT METAL 129

FT METAL 129

FT METAL 169

FT METAL 172

FT METAL 172

FT METAL 175

FT METAL 219

FT METAL 219

SQ SEQUENCE 843 AA; 88861 MW; 318D67C27AAA193 CRC64;


```

RESULT 5
RT17_YEAST STANDARD; PRT: 131 AA.
ID RT17_YEAST
AC P28778;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Mitochondrial 40S ribosomal protein MRP17.
OS MRP17 OR YKL003C.
GN Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93062809; PubMed=1279374;
RA Haffner P.T., Fox T.D.;
RT "Suppression of carboxy-terminal truncations of the yeast
RT mitochondrial mRNA-specific translational activator PET122 by
RT mutations in two new genes, MRP17 and PET127."
RL Mol. Gen. Genet. 235:64-73(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Boyer J., Pascolo S., Richard G.F., Ghazvini M., Colleaux L.,
RA Thierry A., Monnier A.L., Dujon B.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96242157; PubMed=8668135;
RA Hashida-Okado T., Ogawa A., Endo M., Yasumoto R., Takesako K.,
RA Kato I.;
RT "AUR1, a novel gene conferring aureobasidin resistance on
RT Saccharomyces cerevisiae: a study of defective morphologies in
RT Aur1p-depleted cells."
RL Mol. Gen. Genet. 251:236-244(1996).
CC -!- FUNCTION: COMPONENT OF THE SMALL SUBUNIT OF MITOCHONDRIAL
CC RIBOSOME.
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58362; CAA41256.1; -
DR EMBL: Z28003; CAA81835.1; -
DR EMBL: U49090; AAB06941.1; -
DR PIR: S30119; S30119.
DR SGD: S0001486; MRP17.
KM Ribosomal protein; Mitochondrion.
SQ SEQUENCE 131 AA; 15021 MW; FEED82FBEED7A89F CRC64;

```

```

Query Match 0.7%; Score 7; DB 1; Length 131;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1047 NEDYOSI 1053
DB 125 NEDYOSI 131

```

```

RESULT 6
ID RL15_ECOLI STANDARD; PRT: 144 AA.
AC P02413;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE 50S ribosomal protein L15.
GN RPLO OR B3301.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83220807; PubMed=622285;
RA Ceretti D.P., Dean D., Davis G.R., Bedwell D.M., Nomura M.;
RT "The spc ribosomal protein operon of Escherichia coli: sequence and
RT cotranscription of the ribosomal protein genes and a protein export
RT gene."
RL Nucleic Acids Res. 11:2599-2616(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RX Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RX Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RX Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RX Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE.
RC STRAIN=K;
RX MEDLINE=78084799; PubMed=340263;
RX Georgiannis S., Chen R.;
RT "The primary structure of protein L15 located at the
RT peptidyltransferase center of Escherichia coli ribosomes."
RL FEBS Lett. 84:347-350(1977).
RN [4]
RP MASS SPECTROMETRY.
RX MEDLINE=99196679; PubMed=10094780;
RA Arnold R.J., Reilly J.P.;
RT "Observation of Escherichia coli ribosomal proteins and their
RT posttranslational modifications by mass spectrometry."
RL Anal. Biochem. 269:105-112(1999).
CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.
CC -!- MASS SPECTROMETRY: MM=14980.1; METHOD=MALDI.
CC -!- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X01563; CAA25724.1; -
DR EMBL: U18997; AAA58098.1; -
DR EMBL: AE000408; AAC76326.1; -
DR PIR: A02794; R5EC15.
DR Ecogen; EG10876; IPLO.
DR InterPro; IPR001196; Ribosomal_L15.
DR Pfam; PF00256; L15; 1.
DR Pfam; PF01305; Ribosomal_L15; 1.
DR PROSITE; PS00475; RIBOSOMAL_L15; 1.
KM Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 144 AA; 14980 MW; 53D14CD948B15FD9 CRC64;

```

```

Query Match 0.7%; Score 7; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 634 AKVILAG 640
DB 108 AKVILAG 114

```

```

RESULT 7
RL5_HAEIN STANDARD: PRT: 144 AA.
AC P44533:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L15.
CN RPL0 OR RPL15 OR H10797.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Felschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Keriavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ulterback T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Bratton R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32762; AAC2455.1; -
CC TIGR: H10797; -
CC InterPro: IPR001196; Ribosomal_L15.
CC Pfam: PF00256; L15; 1.
CC DR PIR: A24434; A24434.
CC DR PROSITE: PS00475; RIBOSOMAL_L15; 1.
CC KW Ribosomal protein; rRNA-binding; Complete proteome.
CC SEQUENCE 144 AA; 15072 MW; F0CFC80684DC64C5 CRC64;
SQ
Query Match 0.7%; Score 7; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 634 AKVILAG 640
DB 108 AKVILAG 114
RESULT 8
FMO_MORBO STANDARD: PRT: 157 AA.
AC P07640:
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Fimbrial protein Q precursor (beta pilin) (Q pilin).
CN TFPO.
OS Moraxella bovis.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
OX NCBI_TaxID=476;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=EP63;
RX MEDLINE=85234350; PubMed=2861194;
RA Marrs C.F., Schoolnik G., Roomey J.M., Hardy J., Rothbard J.,
RA Falkow S.;
RT "Cloning and sequencing of a Moraxella bovis pilin gene."
RT J. Bacteriol. 163:132-139(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=EP63;
RX MEDLINE=90094235; PubMed=2403542;
RA Fuks K.A., Marrs C.F., Stevens S.P., Green M.R.;
RT "Sequence analysis of the inversion region containing the pilin genes
RT of Moraxella bovis."
RT J. Bacteriol. 172:310-316(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC MEDLINE=91286182; PubMed=2061282;
RX Roza F.W., Marrs C.F.;
RA "Interesting sequence differences between the pilin gene inversion
RT regions of Moraxella lacunata ATCC 17956 and Moraxella bovis EP63."
RL J. Bacteriol. 173:4000-4006(1991).
RN [4]
RP SEQUENCE OF 7-157.
RX MEDLINE=89010522; PubMed=2902184;
RA Ruel W.W., Marrs C.F., Fernandez R., Falkow S., Schoolnik G.K.;
RT "Purification, characterization, and pathogenicity of Moraxella bovis
RT pilin."
RL J. Exp. Med. 168:983-1002(1988).
CC -1- SUBUNIT: THE PILI ARE POLAR FLEXIBLE FILAMENTS OF ABOUT 5.4
CC NANOMETERS DIAMETER AND 2.5 MICROMETERS AVERAGE LENGTH; THEY
CC CONSIST OF ONLY A SINGLE POLYPEPTIDE CHAIN ARRANGED IN A HELICAL
CC CONFIGURATION OF FIVE SUBUNITS PER TURN IN THE ASSEMBLED PILUS.
CC -1- MISCELLANEOUS: MORAXELLA BOVIS CAN EXPRESS EITHER A Q OR A I
CC PILIN, THE INVERSION OF 2 KB OF DNA DETERMINES WHICH PILIN IS
CC EXPRESSED.
CC -----
CC -1- SIMILARITY: BELONGS TO THE N-ME-PHE PILI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M1435; AAA25304.1; -
CC DR EMBL: M32345; AAA88223.1; -
CC DR EMBL: M59712; AAA25308.1; -
CC DR PIR: A24434; A24434.
CC DR PIR: J10072; J10072.
CC DR InterPro: IPR001082; Pilin.
CC DR InterPro: IPR001120; Prok_N_methyltn.
CC Pfam: PF00114; Pilin; 1.
CC DR PRODOM: PD000666; Pilin; 1.
CC DR PROSITE: PS00409; PROKAR_NTER_METHYL; 1.
CC KW Fimbria; Methylation.
CC FT PROPEP 1 6
CC FT CHAIN 7 157
CC FT MOD_RES 7 7
CC FT DISULFID 136 155
CC FT METHYLATION.
CC FT BY SIMILARITY.
SQ SEQUENCE 157 AA; 16006 MW; A923CD8A26C693C9 CRC64;
Query Match 0.7%; Score 7; DB 1; Length 157;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 481 ALFEGK 487
DB 55 ALFEGK 61

```

```

RESULT 9
FMI_MORBO STANDARD; PRT: 159 AA.
AC P20657;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Fimbrial protein I precursor (Alpha pilin) (I pilin).
OS Moraxella bovis.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
ON NCBI_TaxID=476;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=EPB63;
RX MEDLINE=90094235; PubMed=2403542;
RA Fulk's K.A., Marrs C.F., Stevens S.P., Green M.R.;
RT "Sequence analysis of the inversion region containing the pilin genes
of Moraxella bovis.";
RL J. Bacteriol. 172:310-316(1990).
RN
RP SEQUENCE OF 7-159.
RX MEDLINE=9010522; PubMed=2902184;
RA Riehl W.W., Marrs C.F., Fernandez R., Falkow S., Schoolnik G.K.;
RT "Purification, characterization, and pathogenicity of Moraxella bovis
pilin.";
RL J. Exp. Med. 168:983-1002(1988).
CC -1- SUBUNIT: THE PILI ARE POLAR FLEXIBLE FILAMENTS OF ABOUT 5.4
NANOMETERS DIAMETER AND 2.5 MICROMETERS AVERAGE LENGTH; THEY
CONSIST OF ONLY A SINGLE POLYPEPTIDE CHAIN ARRANGED IN A HELICAL
CONFIGURATION OF FIVE SUBUNITS PER TURN IN THE ASSEMBLED PLUS.
CC -1- MISCELLANEOUS: MORAXELLA BOVIS CAN EXPRESS EITHER A O OR A I
PILIN, THE INVERSION OF 2 KB OF DNA DETERMINES WHICH PILIN IS
EXPRESSED.
CC -1- SIMILARITY: BELONGS TO THE N-ME-PHE PILI FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M33345; NOT_ANNOTATED_CDS.
DR PIR: J10071; J10071.
DR HSSP: P02974; IAY2.
DR InterPro: IPR001082; Pilin.
DR InterPro: IPR001120; Prok_N_methyltn.
DR Pfam: PF00114; Pilin; 1.
DR ProDom: PD000666; Pilin; 1.
DR PROSITE: PS00409; PROKAR_NTER_METHYL; 1.
KW Fimbria; Methylation.
FT PROPEP 1 159 FIMBRIAL PROTEIN I.
FT CHAIN 7 159 METHYLATION.
FT MOD_RES 7 159 K->KSK (TN REF. 2).
FT CONFLICT 159 159 16723 MW; 9130E2289C7F679E CRC64;
SQ SEQUENCE 159 AA; 16723 MW; 9130E2289C7F679E CRC64;

Query Match 0.7%; Score 7; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 481 AALFEGK 487
Db 55 AALFEGK 61

RESULT 10
MLE_TODPA STANDARD; PRT: 159 AA.
AC P05945;
DT 01-NOV-1988 (Rel. 09, Created)

```

```

DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin catalytic light chain LC-1, mantle muscle.
OS Todarodes pacificus (Japanese flying squid).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoidea;
OC Oegopsida; Ommastrephidae; Todarodes.
ON NCBI_TaxID=6637;
[1]
RN
RP SEQUENCE.
RX MEDLINE=87076038; PubMed=3790251;
RA Watanabe B., Malta T., Kono K., Matsuda G.;
RT "Amino-acid sequence of LC-1 light chain of squid mantle muscle
myosin.";
RL Biol. Chem. Hoppe-Seyler 367:1025-1032(1986).
CC -1- FUNCTION: IN MOLLUSCAN MUSCLE, CALCIUM REGULATION IS ASSOCIATED
WITH MYOSIN RATHER THAN WITH ACTIN. MUSCLE MYOSIN CONTAINS TWO
TYPES OF LIGHT CHAINS: THE CATALYTIC LIGHT CHAIN, ESSENTIAL FOR
ATPASE ACTIVITY, AND THE REGULATORY LIGHT CHAIN, A CALCIUM-BINDING
PROTEIN RESPONSIBLE FOR CA(++) DEPENDENT BINDING AND CA(++)
DEPENDENT MG-ATPASE ACTIVITY.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS, BUT THIS
PROTEIN DOES NOT BIND CALCIUM.
DR PIR: A25571; A25571.
DR HSSP: P07291; IMDC.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; ethand; 1.
KW Myosin; Muscle protein.
FT MOD_RES 1 1 BLOCKED.
SQ SEQUENCE 159 AA; 18059 MW; 0F60334A34971410 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 923 AKVGDL 929
Db 31 AKVGDL 37

RESULT 11
POPS_YEAST STANDARD; PRT: 173 AA.
AC P28003;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonucleases P/MRP protein subunit POPS (RC 3.1.26.5) (RNases P/MRP
DE 19.6 kDa subunit) (RNA processing protein POPS).
GN POPS OR YAL033W OR FUN53.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
ON NCBI_TaxID=4932;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92260538; PubMed=1583694;
RA Harris S.D., Cheng J., Pugh T.A., Pringle J.R.;
RT "Molecular analysis of Saccharomyces cerevisiae chromosome I. On the
RT number of genes and the identification of essential genes using
RT temperature-sensitive-lethal mutations.";
RL J. Mol. Biol. 225:33-65(1992).
[2]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC -1- FUNCTION: COMPONENT OF RIBONUCLEASE P, A PROTEIN COMPLEX THAT

```



```
CC GENERATES MATURE tRNA MOLECULES BY CLEAVING THEIR 5' ENDS.
CC ALSO A COMPONENT OF RNASE MRP.
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-
CC extra-nucleotide from tRNA precursor.
CC -1- SUBUNIT: COMPONENT OF NUCLEAR RNASE P AND RNASE MRP RNASE P
CC RIBONUCLOPROTEINS. RNASE P CONSISTS OF A RNA MOIETY AND AT LEAST
CC 8 PROTEIN SUBUNITS; POP1, POP3, POP4, POP5, POP6, POP7, POP8 AND
CC RPP1.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X62577; CAA44457.1; -.
CC DR EMBL; U12980; AAC0499.1; -.
CC DR PIR; S23411; S23411.
CC DR SGD; S0000031; POP5.
CC DR InterPro; IPR002759; DUF69.
CC DR Pfam; PF01900; DUF69; 1.
CC DR PRODOM; PD012772; DUF69; 1.
CC KW Hydrolase; Nuclear protein; tRNA processing.
CC SEQUENCE 173 AA; 19573 MW; 918193631BD790DD CRC64;

Query Match 0.7%; Score 7; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1004 LGDYGSA 1010
    |||||
DB 60 LGDYGSA 66

RESULT 12
UCRI_RHORU STANDARD; PRT; 182 AA.
AC P23136;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1989 (Rel. 38, Last annotation update)
DE Ubiquinol-cytochrome c reductase iron-sulfur subunit (EC 1.10.2.2)
DE (Rieske iron-sulfur protein) (RISP).
GN PETA OR FBCE.
OS Rhodospirillum rubrum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC Rhodospirillum.
OX NCBI_TaxID=1085;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-34.
RC STRAIN=FR1;
RX MEDLINE=91094774; PubMed=2176269;
RA Majewski C., Trebst A.;
RT "The pet genes of Rhodospirillum rubrum: cloning and sequencing of
RT the genes for the cytochrome bcl-complex.";
RL Mol. Gen. Genet. 224:373-382(1990).
CC -1- FUNCTION: COMPONENT OF THE UBILQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: QH(2) + 2 ferricytochrome c = Q + 2
CC ferrocytochrome c.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -1- MISCELLANEOUS: THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S
CC PROTEIN.
CC -1- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,
CC BACTERIAL, CHLOROPLAST).
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X55387; CAA39058.1; -.
CC DR PIR; S12256; RDOFBR.
CC DR HSSP; P13272; IRLE.
CC DR InterPro; IPR001281; Rieske.
CC DR Pfam; PF00355; Rieske; 1.
CC DR PRINTS; PR00162; Rieske.
CC DR PROSITE; PS00199; RIESKE_1; 1.
CC DR PROSITE; PS00200; RIESKE_2; 1.
CC KW Electron transport; Inner membrane; Transmembrane; Iron-sulfur;
CC Oxidoreductase.
CC FT INIT_MET 0 0
CC FT TRANSMEM 20 40
CC FT METAL 120 120
CC FT METAL 122 122
CC FT METAL 144 144
CC FT METAL 147 147
CC FT DISULFID 125 146
CC FT BY SIMILARITY.
CC SEQUENCE 182 AA; 19382 MW; 4AC1F24D74CE3A4 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 182;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 526 DRLVYGT 532
    |||||
DB 18 DRLVYGT 24

RESULT 13
YEIP_ECOLI STANDARD; PRT; 190 AA.
ID YEIP_ECOLI
AC P33028;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein yeip.
DE YEIP OR B2171.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robinson K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/Genbank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
```

RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasundaram S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiiuchi T.,
RT "A 460-kb DNA sequence of the *Escherichia coli* K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.",
RL DNA Res. 3:379-392(1996).
RN [1]
RP SEQUENCE OF 50-98 FROM N.A.
RC STRAIN-K12 / W3110;
RA MEDLINE=94162733; PubMed=7764507;
RY Yamada M., Yanai S., Talkner A.;
RT "Analysis of products of the *Escherichia coli* genomic genes and
regulation of their expressions: an applicable procedure for genomic
analysis of other microorganisms.";
RL Biosci. Biotechnol. Blochem. 58:117-120(1994).
RN [5]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=94420866; PubMed=10493123;
RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
RT "Enrichment of low abundance proteins of *Escherichia coli* by
hydroxyapatite chromatography.";
RL Electrophoresis 20:2181-2195(1999).
CC -1- SIMILARITY: BELONGS TO THE ELONGATION FACTOR P FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00007; AAA60519.1; ALT_INIT.
DR EMBL; AE000306; AAC75232.1; ALT_INIT.
DR EMBL; D90849; BAA15980.1; ALT_INIT.
DR EMBL; D21148; BAA04684.1; -
DR Ecocore; EG12035; Yeip.
DR InterPro; IPR001059; EFP.
DR Pfam; PF01132; EFP; 1.
DR PROSITE; PS01275; EFP; 1.
KW Complete proteome.
SQ SEQUENCE 190 AA; 21532 MW; 628EF9E8A26EC7CA CRC64;

Query Match 0.7%; Score 7; DB 1; Length 190;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 DGOALL 376
DB 121 DGOALL 127

RESULT 14
SODE_TETPY STANDARD; PRT; 196 AA.
AC P19666;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Superoxide dismutase [Fe] (EC 1.15.1.1).
OS Tetrahymena pyriformis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5908;
RN [1]
RP SEQUENCE
RX MEDLINE=91009226; PubMed=2170391;
RA Baria D., Schinina M.E., Bossa F., Puget K., Durosay P., Guissani A.,
RA Michelson A.M.;
RT "A tetrahymena iron superoxide dismutase from the eucaryote
Tetrahymena pyriformis.";

RL J. Biol. Chem. 265:17680-17687(1990).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: THE TETRAMER CONTAINS 2.5 G ATOMS OF FERRIC IRON.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
FAMILY.
DR PIR; A39223; A39223.
DR HSSP; P04179; IAP6.
DR InterPro; IPR001189; SOD_MI.
DR Pfam; PF00081; sodefe; 1.
DR Pfam; PF02777; sodefe; 1.
DR ProDom; PD000475; SOD_MI; 1.
DR PROSITE; PS00088; SOD_MN; 1.
KW Oxidoreductase; Iron.
FT METAL 20
FT METAL 68
FT METAL 157
FT METAL 161
SQ SEQUENCE 196 AA; 22657 MW; FC1F2F67893D8C7 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 196;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1060 EKKRYLOA 1066
DB 188 EKKRYLOA 194

RESULT 15
DEFRA_MEDSA STANDARD; PRT; 217 AA.
AC P51109;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Dihydroflavonol-4-reductase (EC 1.1.1.219) (DPR) (Dihydrokaempferol 4-
reductase) (Fragment).
GN DPR1.
OS Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
OX NCBI_TaxID=3879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, VARIA A2;
RX MEDLINE=96128019; PubMed=8541503;
RA Charrier B., Coronado C., Kondorosi A., Ratet P.;
RT "Molecular characterization and expression of alfalfa (*Medicago
sativa* L.) flavanone-3-hydroxylase and dihydroflavonol-4-reductase
encoding genes.";
RL Plant Mol. Biol. 29:773-786(1995).
CC -1- CATALYTIC ACTIVITY: C15-3,4-leucopelargonidin + NADP(+) = (+) -
dihydrokaempferol + NADPH.
CC -1- PATHWAY: FLAVONOID SYNTHESIS; ANTHOCYANIDIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE DIHYDROFLAVONOL-4-REDUCTASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X80222; CAA56508.1; -
RX Flavonoid biosynthesis; Oxidoreductase; NADP.
FT NON_TER 1
FT NON_TER 1
FT NON_TER 217 217

SO SEQUENCE 217 AA; 24389 MM; E8A43AA76CBE3EB2 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 217;

Best Local Similarity 100.0%; Pred. No. 40; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 596 PTIKGVL 602

|||||

Db 85 PTIKGVL 91

Search completed: May 21, 2002, 14:55:21
Job time: 180 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 14:53:21 ; Search time 42.6 Seconds
(without alignments)
4361.421 Million cell updates/sec

Title: US-09-729-653-2_COPY_1_1074
Perfect score: 1074
Sequence: 1 HSLIGRCSRLLDGNVAVC.....LYFEGEKRYLQAGKFLICG 1074

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	length	DB ID	Description
1	498	46.4	905	4	Q9HCD4	Q9HCD4 homo sapien
2	270	25.1	481	4	Q9H5S0	Q9H5S0 homo sapien
3	9	0.8	519	5	O46018	O46018 caenorhabd
4	9	0.8	706	5	O9GY76	O9GY76 leishmania
5	8	0.7	120	15	O995R1	O995R1 human immun
6	8	0.7	176	12	O84636	O84636 paramacium
7	8	0.7	241	17	Q9H087	Q9H087 halobacteri
8	8	0.7	279	16	Q98FH3	Q98FH3 rhizobium l
9	8	0.7	300	2	Q93F73	Q93F73 bacillus sp
10	8	0.7	300	2	Q93F70	Q93F70 bacillus sp
11	8	0.7	474	13	O73906	O73906 gallus gall
12	8	0.7	493	2	O9EV50	O9EV50 staphylococ
13	8	0.7	505	2	O9KM19	O9KM19 staphylococ
14	8	0.7	519	16	Q9CD12	Q9CD12 mycobacteri
15	8	0.7	699	12	O993G4	O993G4 callitrichi
16	8	0.7	1214	10	O91D60	O91D60 arabidopsis

17	8	0.7	1228	5	O61198	O61198 caenorhabd
18	8	0.7	1716	11	O99M7	O99M7 mus musculu
19	8	0.7	1750	5	O9BK11	O9BK11 trypanosoma
20	7	0.7	39	16	Q92TN5	Q92TN5 rhizobium m
21	7	0.7	85	15	O9YUK0	O9YUK0 human immun
22	7	0.7	94	10	O9SK19	O9SK19 arabidopsis
23	7	0.7	98	8	O04290	O04290 phaseolus v
24	7	0.7	98	8	O9MEV2	O9MEV2 sclurus vul
25	7	0.7	100	3	O96U24	O96U24 neurospora
26	7	0.7	105	9	O9B061	O9B061 mycobacteri
27	7	0.7	110	16	Q930C5	Q930C5 rhizobium m
28	7	0.7	118	3	O90Q22	O90Q22 sordaria fi
29	7	0.7	118	3	O90U09	O90U09 sordaria ma
30	7	0.7	118	3	O90U08	O90U08 sordaria sc
31	7	0.7	118	3	O90U04	O90U04 neurospora
32	7	0.7	118	3	O90U04	O90U04 neurospora
33	7	0.7	118	3	O90U02	O90U02 neurospora
34	7	0.7	118	3	O90U08	O90U08 neurospora
35	7	0.7	118	3	O90U06	O90U06 neurospora
36	7	0.7	118	3	O90U01	O90U01 neurospora
37	7	0.7	118	3	O90U00	O90U00 neurospora
38	7	0.7	118	3	O90U00	O90U00 neurospora
39	7	0.7	118	3	O90U03	O90U03 neurospora
40	7	0.7	118	3	O90U06	O90U06 neurospora
41	7	0.7	122	2	O86438	O86438 pseudomonas
42	7	0.7	123	10	O91QV5	O91QV5 arabidopsis
43	7	0.7	126	3	O96U28	O96U28 gelastinospo
44	7	0.7	126	3	O96U27	O96U27 gelastinospo
45	7	0.7	126	3	O96U26	O96U26 gelastinospo
46	7	0.7	126	3	O96U25	O96U25 gelastinospo
47	7	0.7	127	10	O9FTY2	O9FTY2 oryza sativ
48	7	0.7	140	10	O940J7	O940J7 arabidopsis
49	7	0.7	140	16	O9CKR0	O9CKR0 pasteurella
50	7	0.7	141	5	O9VPE7	O9VPE7 drosophila
51	7	0.7	142	5	O94753	O94753 schistosoma
52	7	0.7	142	16	O9CBJ7	O9CBJ7 mycobacteri
53	7	0.7	143	3	O14368	O14368 schizosacch
54	7	0.7	150	2	O9Z120	O9Z120 burkholderi
55	7	0.7	151	12	P88893	P88893 jamaican to
56	7	0.7	153	3	O9HMD7	O9HMD7 paracoccidi
57	7	0.7	153	3	O9HGX6	O9HGX6 paracoccidi
58	7	0.7	153	5	O9YUFA	O9YUFA drosophila
59	7	0.7	154	16	O9CHW9	O9CHW9 lactococcus
60	7	0.7	156	2	O59507	O59507 moraxella b
61	7	0.7	157	2	O9E747	O9E747 bacteroides
62	7	0.7	157	2	O59503	O59503 moraxella b
63	7	0.7	158	2	O59501	O59501 moraxella b
64	7	0.7	158	2	O59508	O59508 moraxella b
65	7	0.7	159	2	O60163	O60163 moraxella b
66	7	0.7	159	2	O59505	O59505 moraxella b
67	7	0.7	160	2	O59504	O59504 moraxella b
68	7	0.7	173	2	O9F4J9	O9F4J9 streptococ
69	7	0.7	179	2	O69605	O69605 methylobact
70	7	0.7	184	8	O63042	O63042 bralithwaite
71	7	0.7	186	10	O9M737	O9M737 phyllocladepa
72	7	0.7	188	8	O63041	O63041 thuidium ta
73	7	0.7	188	10	O9M790	O9M790 trachyloma
74	7	0.7	188	10	O9M730	O9M730 bescherelli
75	7	0.7	189	8	O9BC49	O9BC49 andreaebry
76	7	0.7	189	8	O9BC35	O9BC35 oedipodium
77	7	0.7	189	8	O9BC27	O9BC27 schlottheimi
78	7	0.7	189	8	O9SE40	O9SE40 tayloria ii
79	7	0.7	189	10	O9M7A5	O9M7A5 forstroemi
80	7	0.7	190	2	P96098	P96098 thiodacilli
81	7	0.7	190	10	O9M772	O9M772 thuidium de
82	7	0.7	191	8	O9MUT4	O9MUT4 dawsonia pa
83	7	0.7	191	8	O9GEA2	O9GEA2 campylopus
84	7	0.7	191	8	O9GEA1	O9GEA1 campylopus
85	7	0.7	191	8	O9GE97	O9GE97 dicranodont
86	7	0.7	191	8	O9GGE9	O9GGE9 bescherelli
87	7	0.7	193	10	O9AY09	O9AY09 chamaeabryum
88	7	0.7	194	8	O63029	O63029 tayloria ii
89	7	0.7	201	16	O50930	O50930 borrelia bu

90	7	0.7	201	16	092527	092527 r transposa	163	7	0.7	361	12	09YPS2	09YPS2 vigna mungo
91	7	0.7	207	12	012658	012658 tomato dwar	164	7	0.7	361	12	088549	088549 tomato leaf
92	7	0.7	214	11	063177	063177 ratus norv	165	7	0.7	361	12	091272	091272 tomato leaf
93	7	0.7	214	11	0922H8	0922H8 mus musculu	166	7	0.7	362	12	099DB7	099DB7 vigna mungo
94	7	0.7	214	12	0991R6	0991R6 tomato leaf	167	7	0.7	362	12	091SC1	091SC1 indian mung
95	7	0.7	214	12	0991R4	0991R4 tomato leaf	168	7	0.7	362	12	091P84	091P84 tomato leaf
96	7	0.7	219	16	0921G3	0921G3 rhizobium m	169	7	0.7	362	12	091N70	091N70 mungbean ye
97	7	0.7	223	16	010843	010843 mycobacteri	170	7	0.7	362	12	0913E9	0913E9 indian mung
98	7	0.7	228	12	09WPI2	09WPI2 havana toma	171	7	0.7	362	12	0913E9	0913E9 indian mung
99	7	0.7	229	16	09KSR0	09KSR0 vibrio chol	172	7	0.7	364	12	0910Y7	0910Y7 soybean yel
100	7	0.7	232	10	022922	022922 arabisdopsis	173	7	0.7	364	12	09W7F8	09W7F8 newcastle d
101	7	0.7	232	10	022922	022922 arabisdopsis	174	7	0.7	364	12	09W7F8	09W7F8 newcastle d
102	7	0.7	239	2	09F9K7	09F9K7 pisciticket	175	7	0.7	364	12	088B32	088B32 newcastle d
103	7	0.7	241	16	0989J1	0989J1 rhizobium 1	176	7	0.7	364	12	083833	083833 newcastle d
104	7	0.7	248	17	09V069	09V069 pyrococcus	177	7	0.7	364	12	083835	083835 newcastle d
105	7	0.7	250	16	09KOE5	09KOE5 vibrio chol	178	7	0.7	364	12	083838	083838 newcastle d
106	7	0.7	251	2	09FDR9	09FDR9 microcystis	179	7	0.7	364	12	083839	083839 newcastle d
107	7	0.7	251	2	09RNB4	09RNB4 microcystis	180	7	0.7	364	12	083840	083840 newcastle d
108	7	0.7	252	10	09F120	09F120 arabisdopsis	181	7	0.7	364	12	09WBL3	09WBL3 newcastle d
109	7	0.7	256	17	097CWS	097CWS thermoplasm	182	7	0.7	364	12	090338	090338 newcastle d
110	7	0.7	265	10	039598	039598 chlamydomon	183	7	0.7	364	12	092330	092330 newcastle d
111	7	0.7	266	2	09AKE2	09AKE2 rickettsia	184	7	0.7	364	12	09W6Z2	09W6Z2 newcastle d
112	7	0.7	266	2	09AKO1	09AKO1 rickettsia	185	7	0.7	364	12	09W6Z2	09W6Z2 newcastle d
113	7	0.7	274	5	094849	094849 drosophila	186	7	0.7	364	12	09W6Z4	09W6Z4 newcastle d
114	7	0.7	277	5	094969	094969 drosophila	187	7	0.7	364	12	09W6Z5	09W6Z5 newcastle d
115	7	0.7	282	17	09UWY6	09UWY6 sulfolobus	188	7	0.7	364	12	09W6Z6	09W6Z6 newcastle d
116	7	0.7	285	2	09RL43	09RL43 streptomyc	189	7	0.7	364	12	09W6Z7	09W6Z7 newcastle d
117	7	0.7	292	2	09ACY3	09ACY3 streptomyc	190	7	0.7	364	12	09W6Z8	09W6Z8 newcastle d
118	7	0.7	293	10	09W053	09W053 arabisdopsis	191	7	0.7	364	12	09W6Z9	09W6Z9 newcastle d
119	7	0.7	293	10	093V35	093V35 oryza sativ	192	7	0.7	364	12	09WH00	09WH00 newcastle d
120	7	0.7	295	2	044588	044588 alcaligenes	193	7	0.7	364	12	09WH01	09WH01 newcastle d
121	7	0.7	299	2	093F63	093F63 bacillus sp	194	7	0.7	364	12	09WH02	09WH02 newcastle d
122	7	0.7	305	4	015287	015287 homo sapien	195	7	0.7	364	12	09WH03	09WH03 newcastle d
123	7	0.7	305	4	075308	075308 homo sapien	196	7	0.7	364	12	09DLD5	09DLD5 newcastle d
124	7	0.7	305	11	099M28	099M28 mus musculu	197	7	0.7	364	12	091F52	091F52 newcastle d
125	7	0.7	306	12	091EFT9	091EFT9 cotton leaf	198	7	0.7	364	12	0914X4	0914X4 newcastle d
126	7	0.7	310	2	09FLK3	09FLK3 citrobacter	199	7	0.7	365	17	027876	027876 methanother
127	7	0.7	310	2	09FLK3	09FLK3 citrobacter	200	7	0.7	367	4	096NB9	096NB9 homo sapien
128	7	0.7	310	16	09UOL6	09UOL6 neisseria m	201	7	0.7	371	16	09ZDP9	09ZDP9 rickettsia
129	7	0.7	312	16	067040	067040 aquifex aeo	202	7	0.7	373	4	09H6B4	09H6B4 homo sapien
130	7	0.7	314	2	085929	085929 sphingomon	203	7	0.7	373	11	0920S5	0920S5 mus musculu
131	7	0.7	314	16	092J38	092J38 helicobacte	204	7	0.7	379	16	005775	005775 mycobacteri
132	7	0.7	315	2	09R9L5	09R9L5 rhizobium m	205	7	0.7	388	5	021174	021174 caenorhabdi
133	7	0.7	315	2	09X983	09X983 rhizobium m	206	7	0.7	390	10	093Z35	093Z35 arabisdopsis
134	7	0.7	315	2	052887	052887 rhizobium m	207	7	0.7	391	4	075854	075854 homo sapien
135	7	0.7	315	16	026091	026091 helicobacte	208	7	0.7	392	17	058317	058317 pyrococcus
136	7	0.7	316	16	0986T0	0986T0 rhizobium 1	209	7	0.7	394	16	09X213	09X213 thermotoga
137	7	0.7	318	4	09P0R9	09P0R9 homo sapien	210	7	0.7	396	17	097ZL6	097ZL6 sulfolobus
138	7	0.7	319	4	09P0P6	09P0P6 homo sapien	211	7	0.7	406	2	09X8Y9	09X8Y9 streptomyc
139	7	0.7	322	5	09VEH1	09VEH1 drosophila	212	7	0.7	409	2	09KJG1	09KJG1 streptococ
140	7	0.7	330	5	016721	016721 caenorhabdi	213	7	0.7	412	3	09HDF1	09HDF1 paracoccidi
141	7	0.7	337	10	09GSD9	09GSD9 strongyloce	214	7	0.7	413	3	09HES0	09HES0 paracoccidi
142	7	0.7	337	10	09XG67	09XG67 nicotiana t	215	7	0.7	413	8	09B6C9	09B6C9 yarrowia 11
143	7	0.7	338	2	09RD96	09RD96 streptomyc	216	7	0.7	415	3	060060	060060 schizosacch
144	7	0.7	338	5	091027	091027 caenorhabdi	217	7	0.7	416	3	09H0N9	09H0N9 paracoccidi
145	7	0.7	340	16	092XK1	092XK1 rhizobium m	218	7	0.7	416	3	09HDI9	09HDI9 paracoccidi
146	7	0.7	344	3	078613	078613 emericella	219	7	0.7	416	3	09HES3	09HES3 paracoccidi
147	7	0.7	344	16	071057	071057 bacillus su	220	7	0.7	416	3	09HES2	09HES2 paracoccidi
148	7	0.7	345	10	09SRN7	09SRN7 arabisdopsis	221	7	0.7	416	3	09HES1	09HES1 paracoccidi
149	7	0.7	347	3	09P956	09P956 penicillium	222	7	0.7	416	3	09HER9	09HER9 paracoccidi
150	7	0.7	348	3	078579	078579 emericella	223	7	0.7	416	3	09HER8	09HER8 paracoccidi
151	7	0.7	350	2	043983	043983 acinetobact	224	7	0.7	416	3	09HER7	09HER7 paracoccidi
152	7	0.7	352	5	09XV05	09XV05 caenorhabdi	225	7	0.7	416	3	09HER5	09HER5 paracoccidi
153	7	0.7	353	2	032433	032433 acinetobact	226	7	0.7	416	3	09HER0	09HER0 paracoccidi
154	7	0.7	353	2	052166	052166 pseudomonas	227	7	0.7	425	5	09NFD2	09NFD2 drosophila
155	7	0.7	353	2	052175	052175 pseudomonas	228	7	0.7	425	16	005301	005301 mycobacteri
156	7	0.7	353	3	087210	087210 neurospora	229	7	0.7	427	2	09E280	09E280 brucella ab
157	7	0.7	356	16	099SCH4	099SCH4 staphylococ	230	7	0.7	428	10	09ZPF6	09ZPF6 arabisdopsis
158	7	0.7	358	2	093ED3	093ED3 rhizobium 1	231	7	0.7	431	2	0935K0	0935K0 salmonella
159	7	0.7	359	12	067620	067620 tomato yell	232	7	0.7	431	10	09ZVD8	09ZVD8 arabisdopsis
160	7	0.7	359	12	088949	088949 tomato yell	233	7	0.7	435	5	09H8G7	09H8G7 homo sapien
161	7	0.7	361	12	09YRC0	09YRC0 angled luff	234	7	0.7	435	4	027041	027041 thelleria p
162	7	0.7	361	12	09WH24	09WH24 indian mung	235	7	0.7	435	17	09YEB1	09YEB1 aeropyrum p

382	6	0.6	51	2	Q9F8M3	Q9F8M3 carboxydoth	455	6	0.6	84	15	Q9YUM1	Q9YUM1 human immun
383	6	0.6	51	12	Q9E868	Q9E868 gallid herp	456	6	0.6	84	15	Q9WLE6	Q9WLE6 human immun
384	6	0.6	51	12	Q9E860	Q9E860 gallid herp	457	6	0.6	84	15	Q9YUL9	Q9YUL9 human immun
385	6	0.6	53	2	P77410	P77410 escherichia	458	6	0.6	84	15	Q9YUL6	Q9YUL6 human immun
386	6	0.6	53	5	O19154	O19154 caenorhabdi	459	6	0.6	84	15	Q9PQ01	Q9PQ01 human immun
387	6	0.6	56	11	Q9R082	Q9R082 mus musculu	460	6	0.6	84	16	Q92VP8	Q92VP8 human immun
388	6	0.6	57	2	O05320	O05320 mus musculu	461	6	0.6	85	11	Q88264	Q88264 mus musculu
389	6	0.6	57	12	P89008	P89008 mumps virus	462	6	0.6	85	11	Q9YU06	Q9YU06 human immun
390	6	0.6	57	12	P89010	P89010 mumps virus	463	6	0.6	86	2	Q9REF3	Q9REF3 bradyrhizob
391	6	0.6	57	12	P89013	P89013 mumps virus	464	6	0.6	86	10	Q38768	Q38768 avena sativ
392	6	0.6	57	12	Q98349	Q98349 mumps virus	465	6	0.6	86	12	Q71071	Q71071 bovine aden
393	6	0.6	57	12	Q98350	Q98350 mumps virus	466	6	0.6	86	15	Q9W062	Q9W062 human immun
394	6	0.6	57	12	Q9QJX0	Q9QJX0 mumps virus	467	6	0.6	86	15	Q994N0	Q994N0 human immun
395	6	0.6	57	12	Q9QJW9	Q9QJW9 mumps virus	468	6	0.6	87	10	Q94EVL	Q94EVL oryza sativ
396	6	0.6	57	12	Q9QJW8	Q9QJW8 mumps virus	469	6	0.6	87	13	Q90692	Q90692 gallus gall
397	6	0.6	61	2	Q98351	Q98351 mumps virus	470	6	0.6	88	15	Q9W064	Q9W064 human immun
398	6	0.6	61	2	Q93PR1	Q93PR1 bradyrhizob	471	6	0.6	89	2	Q33M16	Q33M16 azotobacter
399	6	0.6	61	9	Q9AYZ6	Q9AYZ6 bacterioph	472	6	0.6	89	5	Q9VQC5	Q9VQC5 drosophila
400	6	0.6	61	16	Q9JUI4	Q9JUI4 neisseria m	473	6	0.6	89	8	O63701	O63701 nicotiana r
401	6	0.6	64	2	Q9RPL9	Q9RPL9 streptococc	474	6	0.6	89	15	Q98XY4	Q98XY4 human immun
402	6	0.6	64	2	Q9FDN2	Q9FDN2 streptococc	475	6	0.6	89	15	Q9DVP6	Q9DVP6 human immun
403	6	0.6	64	2	Q9F0G5	Q9F0G5 streptococc	476	6	0.6	89	16	Q31715	Q31715 bacillus su
404	6	0.6	64	2	Q9XALA	Q9XALA streptomyc	477	6	0.6	90	16	Q92IX1	Q92IX1 rickettsia
405	6	0.6	65	16	Q92JL5	Q92JL5 rickettsia	478	6	0.6	91	10	Q9LEB2	Q9LEB2 oryza sativ
406	6	0.6	66	10	Q23929	Q23929 flaveria pr	479	6	0.6	91	11	Q9CTE0	Q9CTE0 mus musculu
407	6	0.6	66	10	Q23932	Q23932 flaveria tr	480	6	0.6	91	12	Q9PYT8	Q9PYT8 xestia c-ni
408	6	0.6	66	12	Q9NML7	Q9NML7 human polio	481	6	0.6	91	15	Q9PKX5	Q9PKX5 human immun
409	6	0.6	66	12	Q9NML5	Q9NML5 human polio	482	6	0.6	91	15	Q9PKX4	Q9PKX4 human immun
410	6	0.6	67	5	Q9NM93	Q9NM93 leishmania	483	6	0.6	91	15	Q9Q4U9	Q9Q4U9 human immun
411	6	0.6	68	16	Q32072	Q32072 bacillus su	484	6	0.6	91	15	Q9Q4U5	Q9Q4U5 human immun
412	6	0.6	68	16	Q97NV5	Q97NV5 streptococc	485	6	0.6	91	15	Q9Q4U1	Q9Q4U1 human immun
413	6	0.6	70	7	Q95194	Q95194 stizostedio	486	6	0.6	91	15	Q98XX8	Q98XX8 human immun
414	6	0.6	70	7	Q95192	Q95192 stizostedio	487	6	0.6	91	15	Q98XX2	Q98XX2 human immun
415	6	0.6	70	7	Q95191	Q95191 stizostedio	488	6	0.6	91	16	Q98XX4	Q98XX4 human immun
416	6	0.6	70	7	Q95188	Q95188 stizostedio	489	6	0.6	91	16	Q9PEL4	Q9PEL4 xylella fas
417	6	0.6	70	7	Q95183	Q95183 stizostedio	490	6	0.6	92	2	Q9KVR2	Q9KVR2 vibrio chol
418	6	0.6	71	2	Q9ZFR3	Q9ZFR3 mastigoclad	491	6	0.6	92	15	O51557	O51557 pseudomonas
419	6	0.6	71	4	Q13969	Q13969 homo sapien	492	6	0.6	92	15	Q77705	Q77705 human immun
420	6	0.6	71	16	Q92224	Q92224 rhizobium m	493	6	0.6	92	11	Q77709	Q77709 human immun
421	6	0.6	72	2	Q9AN03	Q9AN03 bradyrhizob	494	6	0.6	93	11	O89082	O89082 mus musculu
422	6	0.6	72	5	Q9XZL5	Q9XZL5 comus stria	495	6	0.6	93	15	Q75613	Q75613 human immun
423	6	0.6	72	17	Q97W34	Q97W34 sulfolobus	496	6	0.6	94	2	Q9KRF1	Q9KRF1 clostridium
424	6	0.6	72	17	Q974G6	Q974G6 sulfolobus	497	6	0.6	94	15	Q91X08	Q91X08 human immun
425	6	0.6	73	2	Q9RJ22	Q9RJ22 streptomyc	498	6	0.6	95	17	Q27927	Q27927 methanother
426	6	0.6	76	16	Q9KAK0	Q9KAK0 bacillus ha	499	6	0.6	96	12	P87596	P87596 cowpox viru
427	6	0.6	77	2	Q9Z4G1	Q9Z4G1 shigella so	500	6	0.6	96	15	Q9DVQ9	Q9DVQ9 human immun
428	6	0.6	77	3	Q96WH0	Q96WH0 verticilliu	501	6	0.6	96	15	Q9DVQ6	Q9DVQ6 human immun
429	6	0.6	77	3	Q96WF6	Q96WF6 torrubella	502	6	0.6	97	10	Q9DVP3	Q9DVP3 human immun
430	6	0.6	77	3	Q96WMA	Q96WMA escovopsis	503	6	0.6	97	15	Q93WA2	Q93WA2 arabidopsis
431	6	0.6	77	3	Q96WA3	Q96WA3 escovopsis	504	6	0.6	97	15	Q9Q4U7	Q9Q4U7 human immun
432	6	0.6	77	4	Q9NS11	Q9NS11 homo sapien	505	6	0.6	97	15	Q9Q4T6	Q9Q4T6 human immun
433	6	0.6	77	16	Q9PGF7	Q9PGF7 xylella fas	506	6	0.6	97	15	Q98XY6	Q98XY6 human immun
434	6	0.6	78	11	Q9QW11	Q9QW11 mus sp. a-r	507	6	0.6	97	16	O51726	O51726 borella bu
435	6	0.6	79	2	Q9AEV9	Q9AEV9 lactococcus	508	6	0.6	98	10	Q9FHV4	Q9FHV4 arabidopsis
436	6	0.6	79	4	Q9H566	Q9H566 homo sapien	509	6	0.6	98	15	Q9YOV8	Q9YOV8 human immun
437	6	0.6	79	12	O68100	O68100 human cytom	510	6	0.6	98	15	Q9YOV4	Q9YOV4 human immun
438	6	0.6	79	15	Q98Y00	Q98Y00 human immun	511	6	0.6	98	15	Q91D06	Q91D06 human immun
439	6	0.6	79	16	O25449	O25449 helicobacte	512	6	0.6	98	15	Q9YD76	Q9YD76 human immun
440	6	0.6	79	16	Q9ZL89	Q9ZL89 helicobacte	513	6	0.6	98	15	Q9Q4V7	Q9Q4V7 human immun
441	6	0.6	80	2	Q9EWN2	Q9EWN2 streptomyc	514	6	0.6	98	15	Q9Q4U4	Q9Q4U4 human immun
442	6	0.6	80	12	Q9YVZ7	Q9YVZ7 melaniplus	515	6	0.6	98	15	Q9Q4T4	Q9Q4T4 human immun
443	6	0.6	80	15	Q9PFX3	Q9PFX3 human immun	516	6	0.6	98	15	Q91X04	Q91X04 human immun
444	6	0.6	81	3	Q9I185	Q9I185 schizosacch	517	6	0.6	99	2	O12935	O12935 mycobacteri
445	6	0.6	81	16	Q9CKG4	Q9CKG4 pasteurella	518	6	0.6	99	8	Q9WJMO	Q9WJMO scenedesmus
446	6	0.6	82	8	Q9TH23	Q9TH23 nicotiana r	519	6	0.6	99	11	Q9CVU7	Q9CVU7 mus musculu
447	6	0.6	82	8	Q9XIX8	Q9XIX8 nicotiana t	520	6	0.6	99	15	O10177	O10177 human immun
448	6	0.6	82	10	Q9ZPW0	Q9ZPW0 arabidopsis	521	6	0.6	99	15	Q9PKX7	Q9PKX7 human immun
449	6	0.6	82	16	Q91625	Q91625 pseudomonas	522	6	0.6	99	15	Q9W8A9	Q9W8A9 human immun
450	6	0.6	83	15	Q9YU08	Q9YU08 human immun	523	6	0.6	99	15	Q9E9W3	Q9E9W3 human immun
451	6	0.6	83	15	Q9YU06	Q9YU06 human immun	524	6	0.6	99	15	Q9E9W1	Q9E9W1 human immun
452	6	0.6	84	12	Q91F32	Q91F32 cydia pomon	525	6	0.6	99	15	Q9E9W0	Q9E9W0 human immun
453	6	0.6	84	15	Q9YUM4	Q9YUM4 human immun	526	6	0.6	99	15	O10764	O10764 human immun
454	6	0.6	84	15	Q9YUM3	Q9YUM3 human immun	527	6	0.6	99	15	O12868	O12868 human immun

528	6	0.6	99	15	010766	010766	human	immun	601	038875	038875	human	immun
529	6	0.6	99	15	010767	010767	human	immun	602	038878	038878	human	immun
530	6	0.6	99	15	012869	012869	human	immun	603	038879	038879	human	immun
531	6	0.6	99	15	010770	010770	human	immun	604	038880	038880	human	immun
532	6	0.6	99	15	010774	010774	human	immun	605	038881	038881	human	immun
533	6	0.6	99	15	010776	010776	human	immun	606	038882	038882	human	immun
534	6	0.6	99	15	010779	010779	human	immun	607	038883	038883	human	immun
535	6	0.6	99	15	010780	010780	human	immun	608	038884	038884	human	immun
536	6	0.6	99	15	012870	012870	human	immun	609	038885	038885	human	immun
537	6	0.6	99	15	036753	036753	human	immun	610	038887	038887	human	immun
538	6	0.6	99	15	036754	036754	human	immun	611	038888	038888	human	immun
539	6	0.6	99	15	038711	038711	human	immun	612	038889	038889	human	immun
540	6	0.6	99	15	038712	038712	human	immun	613	038890	038890	human	immun
541	6	0.6	99	15	038713	038713	human	immun	614	038891	038891	human	immun
542	6	0.6	99	15	038714	038714	human	immun	615	038892	038892	human	immun
543	6	0.6	99	15	038715	038715	human	immun	616	038893	038893	human	immun
544	6	0.6	99	15	038774	038774	human	immun	617	038894	038894	human	immun
545	6	0.6	99	15	038780	038780	human	immun	618	038895	038895	human	immun
546	6	0.6	99	15	038788	038788	human	immun	619	038896	038896	human	immun
547	6	0.6	99	15	038790	038790	human	immun	620	038935	038935	human	immun
548	6	0.6	99	15	038794	038794	human	immun	621	038937	038937	human	immun
549	6	0.6	99	15	038797	038797	human	immun	622	038945	038945	human	immun
550	6	0.6	99	15	038798	038798	human	immun	623	038946	038946	human	immun
551	6	0.6	99	15	038799	038799	human	immun	624	038955	038955	human	immun
552	6	0.6	99	15	038800	038800	human	immun	625	038965	038965	human	immun
553	6	0.6	99	15	038801	038801	human	immun	626	038969	038969	human	immun
554	6	0.6	99	15	038802	038802	human	immun	627	038970	038970	human	immun
555	6	0.6	99	15	038803	038803	human	immun	628	039046	039046	human	immun
556	6	0.6	99	15	038807	038807	human	immun	629	039050	039050	human	immun
557	6	0.6	99	15	038808	038808	human	immun	630	039052	039052	human	immun
558	6	0.6	99	15	038810	038810	human	immun	631	039053	039053	human	immun
559	6	0.6	99	15	038811	038811	human	immun	632	039055	039055	human	immun
560	6	0.6	99	15	038812	038812	human	immun	633	039058	039058	human	immun
561	6	0.6	99	15	038813	038813	human	immun	634	039059	039059	human	immun
562	6	0.6	99	15	038814	038814	human	immun	635	039062	039062	human	immun
563	6	0.6	99	15	038815	038815	human	immun	636	039063	039063	human	immun
564	6	0.6	99	15	038816	038816	human	immun	637	039065	039065	human	immun
565	6	0.6	99	15	038817	038817	human	immun	638	039067	039067	human	immun
566	6	0.6	99	15	038822	038822	human	immun	639	039068	039068	human	immun
567	6	0.6	99	15	038824	038824	human	immun	640	039070	039070	human	immun
568	6	0.6	99	15	038828	038828	human	immun	641	039071	039071	human	immun
569	6	0.6	99	15	038829	038829	human	immun	642	039072	039072	human	immun
570	6	0.6	99	15	038830	038830	human	immun	643	039074	039074	human	immun
571	6	0.6	99	15	038832	038832	human	immun	644	039075	039075	human	immun
572	6	0.6	99	15	038833	038833	human	immun	645	039076	039076	human	immun
573	6	0.6	99	15	038836	038836	human	immun	646	039079	039079	human	immun
574	6	0.6	99	15	038838	038838	human	immun	647	039080	039080	human	immun
575	6	0.6	99	15	038840	038840	human	immun	648	039081	039081	human	immun
576	6	0.6	99	15	038841	038841	human	immun	649	039084	039084	human	immun
577	6	0.6	99	15	038842	038842	human	immun	650	039085	039085	human	immun
578	6	0.6	99	15	038844	038844	human	immun	651	039086	039086	human	immun
579	6	0.6	99	15	038845	038845	human	immun	652	039091	039091	human	immun
580	6	0.6	99	15	038847	038847	human	immun	653	039092	039092	human	immun
581	6	0.6	99	15	038848	038848	human	immun	654	039093	039093	human	immun
582	6	0.6	99	15	038850	038850	human	immun	655	039094	039094	human	immun
583	6	0.6	99	15	038853	038853	human	immun	656	039096	039096	human	immun
584	6	0.6	99	15	038854	038854	human	immun	657	039097	039097	human	immun
585	6	0.6	99	15	038855	038855	human	immun	658	039100	039100	human	immun
586	6	0.6	99	15	038856	038856	human	immun	659	039101	039101	human	immun
587	6	0.6	99	15	038857	038857	human	immun	660	039102	039102	human	immun
588	6	0.6	99	15	038858	038858	human	immun	661	039103	039103	human	immun
589	6	0.6	99	15	038859	038859	human	immun	662	039104	039104	human	immun
590	6	0.6	99	15	038860	038860	human	immun	663	039105	039105	human	immun
591	6	0.6	99	15	038861	038861	human	immun	664	039138	039138	human	immun
592	6	0.6	99	15	038862	038862	human	immun	665	039142	039142	human	immun
593	6	0.6	99	15	038863	038863	human	immun	666	039144	039144	human	immun
594	6	0.6	99	15	038864	038864	human	immun	667	039152	039152	human	immun
595	6	0.6	99	15	038865	038865	human	immun	668	039156	039156	human	immun
596	6	0.6	99	15	038868	038868	human	immun	669	039159	039159	human	immun
597	6	0.6	99	15	038869	038869	human	immun	670	039160	039160	human	immun
598	6	0.6	99	15	038872	038872	human	immun	671	092144	092144	human	immun
599	6	0.6	99	15	038873	038873	human	immun	672	090650	090650	human	immun
600	6	0.6	99	15	038874	038874	human	immun	673	090651	090651	human	immun

674	6	0.6	99	15	090652	090652	human	immun	747	6	0.6	99	15	09DRX7	09dix7	human	immun
675	6	0.6	99	15	090653	090653	human	immun	748	6	0.6	99	15	09DRX6	09dix6	human	immun
676	6	0.6	99	15	090654	090654	human	immun	749	6	0.6	99	15	09IK24	09ik24	human	immun
677	6	0.6	99	15	090655	090655	human	immun	750	6	0.6	99	15	09IK22	09ik22	human	immun
678	6	0.6	99	15	090656	090656	human	immun	751	6	0.6	99	15	09IK21	09ik21	human	immun
679	6	0.6	99	15	090657	090657	human	immun	752	6	0.6	99	15	09IK16	09ik16	human	immun
680	6	0.6	99	15	090659	090659	human	immun	753	6	0.6	99	15	09IJG6	09ijg6	human	immun
681	6	0.6	99	15	090660	090660	human	immun	754	6	0.6	99	15	09IJG5	09ijg5	human	immun
682	6	0.6	99	15	090661	090661	human	immun	755	6	0.6	99	15	09JOC8	09joc8	human	immun
683	6	0.6	99	15	090652	090652	human	immun	756	6	0.6	99	15	09JOC7	09joc7	human	immun
684	6	0.6	99	15	090651	090651	human	immun	757	6	0.6	99	15	09JOC7	09joc7	human	immun
685	6	0.6	99	15	090644	090644	human	immun	758	6	0.6	99	15	09JOC7	09joc7	human	immun
686	6	0.6	99	15	090620	090620	human	immun	759	6	0.6	99	15	09JOC7	09joc7	human	immun
687	6	0.6	99	15	09WRG7	09WRG7	human	immun	760	6	0.6	99	15	09JOC7	09joc7	human	immun
688	6	0.6	99	15	09WRG6	09WRG6	human	immun	761	6	0.6	99	15	09JOC7	09joc7	human	immun
689	6	0.6	99	15	09WCV8	09WCV8	human	immun	762	6	0.6	99	15	09JOC7	09joc7	human	immun
690	6	0.6	99	15	09QCL12	09QCL12	human	immun	763	6	0.6	99	15	09JOC7	09joc7	human	immun
691	6	0.6	99	15	09QCL11	09QCL11	human	immun	764	6	0.6	99	15	09JOC7	09joc7	human	immun
692	6	0.6	99	15	09QCLF3	09QCLF3	human	immun	765	6	0.6	99	15	09JOC7	09joc7	human	immun
693	6	0.6	99	15	09QCLF2	09QCLF2	human	immun	766	6	0.6	99	15	09JOC7	09joc7	human	immun
694	6	0.6	99	15	09QCLF1	09QCLF1	human	immun	767	6	0.6	99	15	09JOC7	09joc7	human	immun
695	6	0.6	99	15	09QCLF0	09QCLF0	human	immun	768	6	0.6	99	15	09JOC7	09joc7	human	immun
696	6	0.6	99	15	09QCLD8	09QCLD8	human	immun	769	6	0.6	99	15	09JOC7	09joc7	human	immun
697	6	0.6	99	15	09QCLD7	09QCLD7	human	immun	770	6	0.6	99	15	09JOC7	09joc7	human	immun
698	6	0.6	99	15	09QCLD6	09QCLD6	human	immun	771	6	0.6	99	15	09JOC7	09joc7	human	immun
699	6	0.6	99	15	09QCLD5	09QCLD5	human	immun	772	6	0.6	99	15	09JOC7	09joc7	human	immun
700	6	0.6	99	15	09QCLD4	09QCLD4	human	immun	773	6	0.6	99	15	09JOC7	09joc7	human	immun
701	6	0.6	99	15	09WQO9	09WQO9	human	immun	774	6	0.6	99	15	09JOC7	09joc7	human	immun
702	6	0.6	99	15	09WQO8	09WQO8	human	immun	775	6	0.6	99	15	09JOC7	09joc7	human	immun
703	6	0.6	99	15	09WQO7	09WQO7	human	immun	776	6	0.6	99	15	09JOC7	09joc7	human	immun
704	6	0.6	99	15	09WQO6	09WQO6	human	immun	777	6	0.6	99	15	09JOC7	09joc7	human	immun
705	6	0.6	99	15	09WQO5	09WQO5	human	immun	778	6	0.6	99	15	09JOC7	09joc7	human	immun
706	6	0.6	99	15	09WQO4	09WQO4	human	immun	779	6	0.6	99	15	09JOC7	09joc7	human	immun
707	6	0.6	99	15	09WQO3	09WQO3	human	immun	780	6	0.6	99	15	09JOC7	09joc7	human	immun
708	6	0.6	99	15	09WQO2	09WQO2	human	immun	781	6	0.6	99	15	09JOC7	09joc7	human	immun
709	6	0.6	99	15	09WQO1	09WQO1	human	immun	782	6	0.6	99	15	09JOC7	09joc7	human	immun
710	6	0.6	99	15	09QIS6	09QIS6	human	immun	783	6	0.6	99	15	09JOC7	09joc7	human	immun
711	6	0.6	99	15	09QIS5	09QIS5	human	immun	784	6	0.6	99	15	09JOC7	09joc7	human	immun
712	6	0.6	99	15	09IM96	09IM96	human	immun	785	6	0.6	99	15	09JOC7	09joc7	human	immun
713	6	0.6	99	15	09EAV9	09EAV9	human	immun	786	6	0.6	99	15	09JOC7	09joc7	human	immun
714	6	0.6	99	15	09E818	09E818	human	immun	787	6	0.6	99	15	09JOC7	09joc7	human	immun
715	6	0.6	99	15	09E7Y7	09E7Y7	human	immun	788	6	0.6	99	15	09JOC7	09joc7	human	immun
716	6	0.6	99	15	09E7X8	09E7X8	human	immun	789	6	0.6	99	15	09JOC7	09joc7	human	immun
717	6	0.6	99	15	09E7X2	09E7X2	human	immun	790	6	0.6	99	15	09JOC7	09joc7	human	immun
718	6	0.6	99	15	09E7X1	09E7X1	human	immun	791	6	0.6	99	15	09JOC7	09joc7	human	immun
719	6	0.6	99	15	09E7U9	09E7U9	human	immun	792	6	0.6	99	15	09JOC7	09joc7	human	immun
720	6	0.6	99	15	09E7U8	09E7U8	human	immun	793	6	0.6	99	15	09JOC7	09joc7	human	immun
721	6	0.6	99	15	09E7U5	09E7U5	human	immun	794	6	0.6	99	15	09JOC7	09joc7	human	immun
722	6	0.6	99	15	09E7U4	09E7U4	human	immun	795	6	0.6	99	15	09JOC7	09joc7	human	immun
723	6	0.6	99	15	09E7U1	09E7U1	human	immun	796	6	0.6	99	15	09JOC7	09joc7	human	immun
724	6	0.6	99	15	09Q4W3	09Q4W3	human	immun	797	6	0.6	99	15	09JOC7	09joc7	human	immun
725	6	0.6	99	15	09Q4W1	09Q4W1	human	immun	798	6	0.6	99	15	09JOC7	09joc7	human	immun
726	6	0.6	99	15	09Q4V4	09Q4V4	human	immun	799	6	0.6	99	15	09JOC7	09joc7	human	immun
727	6	0.6	99	15	09Q4V3	09Q4V3	human	immun	800	6	0.6	99	15	09JOC7	09joc7	human	immun
728	6	0.6	99	15	09Q4V0	09Q4V0	human	immun	801	6	0.6	99	15	09JOC7	09joc7	human	immun
729	6	0.6	99	15	09Q4U8	09Q4U8	human	immun	802	6	0.6	99	15	09JOC7	09joc7	human	immun
730	6	0.6	99	15	09Q4U3	09Q4U3	human	immun	803	6	0.6	99	15	09JOC7	09joc7	human	immun
731	6	0.6	99	15	09Q4U2	09Q4U2	human	immun	804	6	0.6	99	15	09JOC7	09joc7	human	immun
732	6	0.6	99	15	09Q4T8	09Q4T8	human	immun	805	6	0.6	99	15	09JOC7	09joc7	human	immun
733	6	0.6	99	15	09ILX1	09ILX1	human	immun	806	6	0.6	99	15	09JOC7	09joc7	human	immun
734	6	0.6	99	15	09ILX1	09ILX1	human	immun	807	6	0.6	99	15	09JOC7	09joc7	human	immun
735	6	0.6	99	15	09ILF7	09ILF7	human	immun	808	6	0.6	99	15	09JOC7	09joc7	human	immun
736	6	0.6	99	15	09ILF1	09ILF1	human	immun	809	6	0.6	99	15	09JOC7	09joc7	human	immun
737	6	0.6	99	15	09ILP5	09ILP5	human	immun	810	6	0.6	99	15	09JOC7	09joc7	human	immun
738	6	0.6	99	15	09ILP1	09ILP1	human	immun	811	6	0.6	99	15	09JOC7	09joc7	human	immun
739	6	0.6	99	15	09ILN8	09ILN8	human	immun	812	6	0.6	99	15	09JOC7	09joc7	human	immun
740	6	0.6	99	15	09J2W5	09J2W5	human	immun	813	6	0.6	99	15	09JOC7	09joc7	human	immun
741	6	0.6	99	15	09J2V9	09J2V9	human	immun	814	6	0.6	99	15	09JOC7	09joc7	human	immun
742	6	0.6	99	15	09J2S2	09J2S2	human	immun	815	6	0.6	99	15	09JOC7	09joc7	human	immun
743	6	0.6	99	15	09J2S2	09J2S2	human	immun	816	6	0.6	99	15	09JOC7	09joc7	human	immun
744	6	0.6	99	15	09J2N2	09J2N2	human	immun	817	6	0.6	99	15	09JOC7	09joc7	human	immun
745	6	0.6	99	15	09DRY0	09DRY0	human	immun	818	6	0.6	99	15	09JOC7	09joc7	human	immun
746	6	0.6	99	15	09DRX8	09DRX8	human	immun	819	6	0.6	99	15	09JOC7	09joc7	human	immun

820	6	0.6	99	15	Q9DMG1	Q9dmg1 human	immun	893	6	0.6	99	15	Q9DM58	Q9dm58 human	immun
821	6	0.6	99	15	Q9DMG0	Q9dmg0 human	immun	894	6	0.6	99	15	Q9DM57	Q9dm57 human	immun
822	6	0.6	99	15	Q9DMF9	Q9dmf9 human	immun	895	6	0.6	99	15	Q9DM56	Q9dm56 human	immun
823	6	0.6	99	15	Q9DMF8	Q9dmf8 human	immun	896	6	0.6	99	15	Q9DM55	Q9dm55 human	immun
824	6	0.6	99	15	Q9DMF7	Q9dmf7 human	immun	897	6	0.6	99	15	Q9DM54	Q9dm54 human	immun
825	6	0.6	99	15	Q9DMF6	Q9dmf6 human	immun	898	6	0.6	99	15	Q9DM53	Q9dm53 human	immun
826	6	0.6	99	15	Q9DMF5	Q9dmf5 human	immun	899	6	0.6	99	15	Q9DM52	Q9dm52 human	immun
827	6	0.6	99	15	Q9DMF4	Q9dmf4 human	immun	900	6	0.6	99	15	Q9DM51	Q9dm51 human	immun
828	6	0.6	99	15	Q9DMF3	Q9dmf3 human	immun	901	6	0.6	99	15	Q9DM50	Q9dm50 human	immun
829	6	0.6	99	15	Q9DMF2	Q9dmf2 human	immun	902	6	0.6	99	15	Q9DM49	Q9dm49 human	immun
830	6	0.6	99	15	Q9DMF1	Q9dmf1 human	immun	903	6	0.6	99	15	Q9DM48	Q9dm48 human	immun
831	6	0.6	99	15	Q9DME9	Q9dme9 human	immun	904	6	0.6	99	15	Q9DM47	Q9dm47 human	immun
832	6	0.6	99	15	Q9DME8	Q9dme8 human	immun	905	6	0.6	99	15	Q9DM46	Q9dm46 human	immun
833	6	0.6	99	15	Q9DME7	Q9dme7 human	immun	906	6	0.6	99	15	Q9DM45	Q9dm45 human	immun
834	6	0.6	99	15	Q9DME6	Q9dme6 human	immun	907	6	0.6	99	15	Q9DM44	Q9dm44 human	immun
835	6	0.6	99	15	Q9DME5	Q9dme5 human	immun	908	6	0.6	99	15	Q9DM43	Q9dm43 human	immun
836	6	0.6	99	15	Q9DME4	Q9dme4 human	immun	909	6	0.6	99	15	Q9DM42	Q9dm42 human	immun
837	6	0.6	99	15	Q9DME2	Q9dme2 human	immun	910	6	0.6	99	15	Q9DM41	Q9dm41 human	immun
838	6	0.6	99	15	Q9DME1	Q9dme1 human	immun	911	6	0.6	99	15	Q9DM40	Q9dm40 human	immun
839	6	0.6	99	15	Q9DMD9	Q9dmd9 human	immun	912	6	0.6	99	15	Q9DM39	Q9dm39 human	immun
840	6	0.6	99	15	Q9DMD8	Q9dmd8 human	immun	913	6	0.6	99	15	Q9DM38	Q9dm38 human	immun
841	6	0.6	99	15	Q9DMD7	Q9dmd7 human	immun	914	6	0.6	99	15	Q9DM37	Q9dm37 human	immun
842	6	0.6	99	15	Q9DMD6	Q9dmd6 human	immun	915	6	0.6	99	15	Q9DM36	Q9dm36 human	immun
843	6	0.6	99	15	Q9DMD5	Q9dmd5 human	immun	916	6	0.6	99	15	Q9DM35	Q9dm35 human	immun
844	6	0.6	99	15	Q9DMD4	Q9dmd4 human	immun	917	6	0.6	99	15	Q9DM34	Q9dm34 human	immun
845	6	0.6	99	15	Q9DMD3	Q9dmd3 human	immun	918	6	0.6	99	15	Q9DM33	Q9dm33 human	immun
846	6	0.6	99	15	Q9DMD2	Q9dmd2 human	immun	919	6	0.6	99	15	Q9DM32	Q9dm32 human	immun
847	6	0.6	99	15	Q9DMD1	Q9dmd1 human	immun	920	6	0.6	99	15	Q9DM31	Q9dm31 human	immun
848	6	0.6	99	15	Q9DMC9	Q9dmc9 human	immun	921	6	0.6	99	15	Q9DM30	Q9dm30 human	immun
849	6	0.6	99	15	Q9DMC8	Q9dmc8 human	immun	922	6	0.6	99	15	Q9DM29	Q9dm29 human	immun
850	6	0.6	99	15	Q9DMC7	Q9dmc7 human	immun	923	6	0.6	99	15	Q9DM28	Q9dm28 human	immun
851	6	0.6	99	15	Q9DMC6	Q9dmc6 human	immun	924	6	0.6	99	15	Q9DM27	Q9dm27 human	immun
852	6	0.6	99	15	Q9DMC5	Q9dmc5 human	immun	925	6	0.6	99	15	Q9DM26	Q9dm26 human	immun
853	6	0.6	99	15	Q9DMC4	Q9dmc4 human	immun	926	6	0.6	99	15	Q9DM25	Q9dm25 human	immun
854	6	0.6	99	15	Q9DMC3	Q9dmc3 human	immun	927	6	0.6	99	15	Q9DM24	Q9dm24 human	immun
855	6	0.6	99	15	Q9DMC2	Q9dmc2 human	immun	928	6	0.6	99	15	Q9DM23	Q9dm23 human	immun
856	6	0.6	99	15	Q9DMC1	Q9dmc1 human	immun	929	6	0.6	99	15	Q9DM22	Q9dm22 human	immun
857	6	0.6	99	15	Q9DMC0	Q9dmc0 human	immun	930	6	0.6	99	15	Q9DM21	Q9dm21 human	immun
858	6	0.6	99	15	Q9DMB8	Q9dmb8 human	immun	931	6	0.6	99	15	Q9DM20	Q9dm20 human	immun
859	6	0.6	99	15	Q9DMB7	Q9dmb7 human	immun	932	6	0.6	99	15	Q9DM19	Q9dm19 human	immun
860	6	0.6	99	15	Q9DMB5	Q9dmb5 human	immun	933	6	0.6	99	15	Q9DM18	Q9dm18 human	immun
861	6	0.6	99	15	Q9DMB4	Q9dmb4 human	immun	934	6	0.6	99	15	Q9DM17	Q9dm17 human	immun
862	6	0.6	99	15	Q9DMB3	Q9dmb3 human	immun	935	6	0.6	99	15	Q9DM16	Q9dm16 human	immun
863	6	0.6	99	15	Q9DMB2	Q9dmb2 human	immun	936	6	0.6	99	15	Q9DM15	Q9dm15 human	immun
864	6	0.6	99	15	Q9DMB1	Q9dmb1 human	immun	937	6	0.6	99	15	Q9DM14	Q9dm14 human	immun
865	6	0.6	99	15	Q9DMB0	Q9dmb0 human	immun	938	6	0.6	99	15	Q9DM13	Q9dm13 human	immun
866	6	0.6	99	15	Q9DMA9	Q9dma9 human	immun	939	6	0.6	99	15	Q9DM12	Q9dm12 human	immun
867	6	0.6	99	15	Q9DMA8	Q9dma8 human	immun	940	6	0.6	99	15	Q9DM11	Q9dm11 human	immun
868	6	0.6	99	15	Q9DMA7	Q9dma7 human	immun	941	6	0.6	99	15	Q9DM10	Q9dm10 human	immun
869	6	0.6	99	15	Q9DMA6	Q9dma6 human	immun	942	6	0.6	99	15	Q9DM09	Q9dm09 human	immun
870	6	0.6	99	15	Q9DMA5	Q9dma5 human	immun	943	6	0.6	99	15	Q9DM08	Q9dm08 human	immun
871	6	0.6	99	15	Q9DMA4	Q9dma4 human	immun	944	6	0.6	99	15	Q9DM07	Q9dm07 human	immun
872	6	0.6	99	15	Q9DMA2	Q9dma2 human	immun	945	6	0.6	99	15	Q9DM06	Q9dm06 human	immun
873	6	0.6	99	15	Q9DMA1	Q9dma1 human	immun	946	6	0.6	99	15	Q9DM05	Q9dm05 human	immun
874	6	0.6	99	15	Q9DMA0	Q9dma0 human	immun	947	6	0.6	99	15	Q9DM04	Q9dm04 human	immun
875	6	0.6	99	15	Q9DM99	Q9dm99 human	immun	948	6	0.6	99	15	Q9DM03	Q9dm03 human	immun
876	6	0.6	99	15	Q9DM98	Q9dm98 human	immun	949	6	0.6	99	15	Q9DM02	Q9dm02 human	immun
877	6	0.6	99	15	Q9DM97	Q9dm97 human	immun	950	6	0.6	99	15	Q9DM01	Q9dm01 human	immun
878	6	0.6	99	15	Q9DM96	Q9dm96 human	immun	951	6	0.6	99	15	Q9DM00	Q9dm00 human	immun
879	6	0.6	99	15	Q9DM95	Q9dm95 human	immun	952	6	0.6	99	15	Q9DL29	Q9dl29 human	immun
880	6	0.6	99	15	Q9DM94	Q9dm94 human	immun	953	6	0.6	99	15	Q9DL28	Q9dl28 human	immun
881	6	0.6	99	15	Q9DM93	Q9dm93 human	immun	954	6	0.6	99	15	Q9DL26	Q9dl26 human	immun
882	6	0.6	99	15	Q9DM92	Q9dm92 human	immun	955	6	0.6	99	15	Q9DL25	Q9dl25 human	immun
883	6	0.6	99	15	Q9DM91	Q9dm91 human	immun	956	6	0.6	99	15	Q9DL24	Q9dl24 human	immun
884	6	0.6	99	15	Q9DM90	Q9dm90 human	immun	957	6	0.6	99	15	Q9DL23	Q9dl23 human	immun
885	6	0.6	99	15	Q9DM66	Q9dm66 human	immun	958	6	0.6	99	15	Q9DL22	Q9dl22 human	immun
886	6	0.6	99	15	Q9DM65	Q9dm65 human	immun	959	6	0.6	99	15	Q9DL21	Q9dl21 human	immun
887	6	0.6	99	15	Q9DM64	Q9dm64 human	immun	960	6	0.6	99	15	Q9DL20	Q9dl20 human	immun
888	6	0.6	99	15	Q9DM63	Q9dm63 human	immun	961	6	0.6	99	15	Q9DL19	Q9dl19 human	immun
889	6	0.6	99	15	Q9DM62	Q9dm62 human	immun	962	6	0.6	99	15	Q9DL18	Q9dl18 human	immun
890	6	0.6	99	15	Q9DM61	Q9dm61 human	immun	963	6	0.6	99	15	Q9DL17	Q9dl17 human	immun
891	6	0.6	99	15	Q9DM60	Q9dm60 human	immun	964	6	0.6	99	15	Q9DL16	Q9dl16 human	immun
892	6	0.6	99	15	Q9DM59	Q9dm59 human	immun	965	6	0.6	99	15	Q9DL15	Q9dl15 human	immun

```

966 6 0.6 99 15 Q9DLX4 Q9DLX4 human immun
967 6 0.6 99 15 Q9DLX3 Q9DLX3 human immun
968 6 0.6 99 15 Q9DLX2 Q9DLX2 human immun
969 6 0.6 99 15 Q9DLX1 Q9DLX1 human immun
970 6 0.6 99 15 Q9DLX0 Q9DLX0 human immun
971 6 0.6 99 15 Q9DLX9 Q9DLX9 human immun
972 6 0.6 99 15 Q9DLX8 Q9DLX8 human immun
973 6 0.6 99 15 Q9DLX7 Q9DLX7 human immun
974 6 0.6 99 15 Q9DLX6 Q9DLX6 human immun
975 6 0.6 99 15 Q9DLX5 Q9DLX5 human immun
976 6 0.6 99 15 Q9DLX4 Q9DLX4 human immun
977 6 0.6 99 15 Q9DLX3 Q9DLX3 human immun
978 6 0.6 99 15 Q9DLX2 Q9DLX2 human immun
979 6 0.6 99 15 Q9DLX1 Q9DLX1 human immun
980 6 0.6 99 15 Q9DLX0 Q9DLX0 human immun
981 6 0.6 99 15 Q9DLW9 Q9DLW9 human immun
982 6 0.6 99 15 Q9DLW8 Q9DLW8 human immun
983 6 0.6 99 15 Q9DLW7 Q9DLW7 human immun
984 6 0.6 99 15 Q9DLW6 Q9DLW6 human immun
985 6 0.6 99 15 Q9DLW5 Q9DLW5 human immun
986 6 0.6 99 15 Q9DLW4 Q9DLW4 human immun
987 6 0.6 99 15 Q9DLW3 Q9DLW3 human immun
988 6 0.6 99 15 Q9DLW2 Q9DLW2 human immun
989 6 0.6 99 15 Q9DLW1 Q9DLW1 human immun
990 6 0.6 99 15 Q9DLW0 Q9DLW0 human immun
991 6 0.6 99 15 Q9DLV9 Q9DLV9 human immun
992 6 0.6 99 15 Q9DLV8 Q9DLV8 human immun
993 6 0.6 99 15 Q9DLV7 Q9DLV7 human immun
994 6 0.6 99 15 Q9DLV6 Q9DLV6 human immun
995 6 0.6 99 15 Q9DLV5 Q9DLV5 human immun
996 6 0.6 99 15 Q9DLV4 Q9DLV4 human immun
997 6 0.6 99 15 Q9DLV3 Q9DLV3 human immun
998 6 0.6 99 15 Q9DLV2 Q9DLV2 human immun
999 6 0.6 99 15 Q9DLV1 Q9DLV1 human immun
1000 6 0.6 99 15 Q9DLU9 Q9DLU9 human immun

```

ALIGNMENTS

```

RESULT 1
ID Q9HCD4 PRELIMINARY: PRT: 905 AA.
AC Q9HCD4:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE KIA1638 PROTEIN (FRAGMENT).
CN KIA1638.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20450683; PubMed-10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;
RT "Prediction of the coding regions of unidentified human genes.
RT XVII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046858; BAB13464.1;
DR InterPro; IPR000547; Clathrin_repeat.
DR SMART; SM00299; CLH; 1.
FT NON_TER 1
SQ SEQUENCE 905 AA: 102450 MW: PD370996F00FE1BF CRC64;

```

```

Query Match 46.4%; Score 498; DB 4; Length 905;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 698; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

OY 375 ALSTORGS.LHVF.LTKLP.LITGDACSTR.IAY.LTSLLEVTVAN.PVEGLP.TIVSV.DVEPNFVA 434

```

|||||
Db 1 ALSTORGS.LHVF.LTKLP.LITGDACSTR.IAY.LTSLLEVTVAN.PVEGLP.TIVSV.DVEPNFVA 60
435 VGLYLAVGMNRAWEYVUGENAVKLLKDMETLGVVASICLHSDVAALFEGKVOHLIE 494
61 VGLYLAVGMNRAWEYVUGENAVKLLKDMETLGVVASICLHSDVAALFEGKVOHLIE 120
495 SEILDAGEERETRLP.PAVDDCKRIICHA.LTSDFL.IYGDGVVOYVEIEMQFVNDYBNP 554
121 SEILDAGEERETRLP.PAVDDCKRIICHA.LTSDFL.IYGDGVVOYVEIEMQFVNDYBNP 180
555 VSVKRIIPDPNKTRLV.FIDEKSDGFVYCPVNDATYEIFDFSTFIVGVLMENPMKGVFI 614
181 VSVKRIIPDPNKTRLV.FIDEKSDGFVYCPVNDATYEIFDFSTFIVGVLMENPMKGVFI 240
615 AYDDDKVTTYVPHKDTIGAKVILAGSTKVPFAHPLLYNGLTCOTQSGKVNNTYST 674
241 AYDDDKVTTYVPHKDTIGAKVILAGSTKVPFAHPLLYNGLTCOTQSGKVNNTYST 300
675 HGFLSLNDXGPDDEL.RPMLAHNMLK.RPSDAMEMCR.IINDEAANMLARACLHMEVEFA 734
301 HGFLSLNDXGPDDEL.RPMLAHNMLK.RPSDAMEMCR.IINDEAANMLARACLHMEVEFA 360
735 IIVYRIRIGNVGIVMSLEQIKGIEDYNLLAGHLAMFTNDYNAQDIYLLASCPIALLEMR 794
361 IIVYRIRIGNVGIVMSLEQIKGIEDYNLLAGHLAMFTNDYNAQDIYLLASCPIALLEMR 420
795 DLQHMDSALQLAKHLAPQIIPISKEVAIQLEFAGDYVNALAHYKGTGDKREDECL 854
421 DLQHMDSALQLAKHLAPQIIPISKEVAIQLEFAGDYVNALAHYKGTGDKREDECL 480
855 AGVAOMSTIRMGDIRRGVNOALKHP.SRVLLK.RDGCALLENKOFSEAAQLYERGLYDKAAS 914
481 AGVAOMSTIRMGDIRRGVNOALKHP.SRVLLK.RDGCALLENKOFSEAAQLYERGLYDKAAS 540
915 VYIRSKNAKVGDDLPHVSSPKIHLOYAKARADRYEAVAYANAKOMOSVIRIYD 974
541 VYIRSKNAKVGDDLPHVSSPKIHLOYAKARADRYEAVAYANAKOMOSVIRIYD 600
975 LNNPEKAVIVRETO.SLDGAKVAVARFLOLDYGSALIOFLVMSKCNNEFTLQOHNKME 1034
601 LNNPEKAVIVRETO.SLDGAKVAVARFLOLDYGSALIOFLVMSKCNNEFTLQOHNKME 660
OY 1035 IVADITIGSEDTTNE.DYOSIALYFEGEKRYLQAKFELLCG 1074
661 IVADITIGSEDTTNE.DYOSIALYFEGEKRYLQAKFELLCG 700
Db
RESULT 2
ID Q9H5S0 PRELIMINARY: PRT: 481 AA.
AC Q9H5S0:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE CDNA: FLJ23127 F1S, CLONE LNC08350.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-LONG;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Kawajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hirose M., Ohmori Y., Ota T., Suzuki Y.,
RA Odayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 3 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL; AK026780; BAB15550.1;
DR InterPro; IPR001680; WD40.

```

DR Pfam: PF00400; WD40; 3.
DR SMART: SM00320; WD40; 5.
KM Repeat: WD repeat.
SQ SEQUENCE 481 AA; 53542 MW; B1C7AE099C581F02 CRC64;

Query Match

25.1%; Score 270; DB 4; Length 481;

Best Local Similarity 100.0%; Pred. No. 3e-288;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 NM0FFLKKMDRTSAASMSISVVGKTTFLFLNLEPDNPADEFGQDFGNIYCYMYGD 282
DB 183 NM0FFLKKMDRTSAASMSISVVGKTTFLFLNLEPDNPADEFGQDFGNIYCYMYGD 242
OY 283 GR1IGSCGFVYISHTGELGEOIFQARNHKNLTSIVASQTLNKAACGDCICITOD 342
DB 243 GR1IGSCGFVYISHTGELGEOIFQARNHKNLTSIVASQTLNKAACGDCICITOD 302
OY 343 LVDLKDMVILNLEENKGLGTLSDGQLALSTORGSLHVEFLTKPLIGDASTRIA 402
DB 303 LVDLKDMVILNLEENKGLGTLSDGQLALSTORGSLHVEFLTKPLIGDASTRIA 362
OY 403 YLTSLEEVYANPEVGLPTVSVDPEPNFYANGYLHLAGMNNRANFYVGENAVKKLK 462
DB 363 YLTSLEEVYANPEVGLPTVSVDPEPNFYANGYLHLAGMNNRANFYVGENAVKKLK 422
OY 463 DMEYLGTVASICLHSDYAALFEGKVOLHL 492
DB 423 DMEYLGTVASICLHSDYAALFEGKVOLHL 452

RESULT 3

ID 046018 PRELIMINARY; PRT; 519 AA.

AC 046018;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ZK520.1 PROTEIN.
CN ZK520.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Steward C.A.;
RN Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX MEDLINE=90659613; Pubmed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology."
DR EMBL: Z92822; CAB07299.1; -;
SQ SEQUENCE 519 AA; 59152 MW; 38F80CA32D06E02D CRC64;

Query Match 0.8%; Score 9; DB 5; Length 519;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1009 SAIOFLVWS 1017
DB 293 SAIOFLVWS 301

RESULT 4
OY 09GY76 PRELIMINARY; PRT; 706 AA.
AC 09GY76;
DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROBABLE HYPOTHETICAL 11.6 KDA PROTEIN.
GN LM12.54.
OS Leishmania major.
OC Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=FRIDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL390114; CAC01977.2; -;
SQ SEQUENCE 706 AA; 72716 MW; 29199F2AA4CC64FD CRC64;

OY 368 TDDGQLLAL 376
DB 388 TDDGQLLAL 396

Query Match 0.8%; Score 9; DB 5; Length 706;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

ID 0995R1 PRELIMINARY; PRT; 120 AA.

AC 0995R1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE . GAG POLYPEPTIDE (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=M6;
RA Yairin D.L.;
RT "Immunological and virological characteristics of HIV infected
individuals in different risk groups in Madras, India."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF250278; AK3492.1; -;
DR InterPro: IPR000071; Retroviral_gag_p17.
DR Pfam: PF00540; gag_p17; 1.
DR PRINTS: PR00234; HIVMATRIX.
KW AIDS; Core protein; Polypeptide.
FT NON_TER 1
FT NON_TER 120
SQ SEQUENCE 120 AA; 13544 MW; 6D4B7597A6ACFB3C CRC64;

Query Match 0.7%; Score 8; DB 15; Length 120;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 942 AKAKEADG 949
DB 90 AKAKEADG 97

RESULT 6
OY 084636 PRELIMINARY; PRT; 176 AA.
AC 084636;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE A322L PROTEIN.
GN A322L.

OS Paramesicium bursaria chlorella virus 1 (PBCV-1).
OC Viruses: dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
OX NCBI_TaxID=10506;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96187795; PubMed=8614977;
RA Lu Z., Li Y., Que Q., Kutish G.F., Rock D.L., Van Etten J.L.;
RT "Analysis of 94 kb of the chlorella virus PBCV-1 330-Kb genome: map
RL positions 88 to 182.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20013326; PubMed=10544099;
RA Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,
RA Liseac A.D., Mickerson K.W., Van Etten J.L.;
RT "Chlorella virus PBCV-1 encodes a functional homospermidine
RN synthase.";
RL Virology 263:254-262(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20478054; PubMed=11021991;
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus
RN PBCV-1.";
RL Virology 276:27-36(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RA Gurnon J.R., Graves M.V., Van Etten J.L.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: U42580; AAC36690.1; -;
DR InterPro: IPR003015; HLH_MYC.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ SEQUENCE 176 AA; 20039 MW; 8CA9FBC4C16A2F96 CRC64;

Query Match 0.7%; Score 8; DB 12; Length 176;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 628 KDTIOGAK 635
DB 103 KDTIOGAK 110

RESULT 7
ID 09HQ87 PRELIMINARY; PRT; 241 AA.
AC 09HQ87;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
GN VNG1276C.

OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasly S.R., Baliga N.S., Thorsson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angewine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RN [2]
RP EMBL: AE005051; MAG19629.1; -;
DR InterPro: IPR002140; UPF0023.
DR Pfam: PF01172; UPF0023; 1.
DR ProDom: PD009796; UPF0023; 1.
KW Complete proteome.

QY 788 AALEMRD 795
DB 27 AALEMRD 34

Query Match 0.7%; Score 8; DB 17; Length 241;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 788 AALEMRD 795
DB 27 AALEMRD 34

RESULT 8
ID 098FH3 PRELIMINARY; PRT; 279 AA.
AC 098FH3;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE STREPTOTHRICIN-ACTEYL-TRANSFERASE.
GN ML3776.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RN Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003002; BAB50594.1; -;
DR InterPro: IPR000073; Abhydrolase.
DR InterPro: IPR003089; AB_hydrolase.
DR InterPro: IPR000379; Est_lip_chloest_actsite.
DR Pfam: PF00561; abhydrolase; 1.
DR PRINTS: PR00111; ABHYDROLASE.
KW Transferase; Complete proteome.

QY SEQUENCE 279 AA; 30294 MW; 8209435C294073A4 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 279;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 874 ALKHSRV 881
 |||||
 DB 109 ALKHSRV 116

RESULT 9
 093F73 PRELIMINARY; PRT; 300 AA.
 AC 093F73;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE MNXG (FRAGMENT).
 GN MNXG.
 OS Bacillus sp. MB-1.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=161530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MB-1;
 RA Francis C.A., Tebo B.M.;
 RT "Enzymatic manganese(II) oxidation by metabolically-dormant spore of
 RT diverse Bacillus species.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF326386; AAL30438.1; -.
 FT NON_TER 1
 FT NON_TER 300
 SQ SEQUENCE 300 AA; 33592 MW; A05B5CA3801F872 CRC64;

Query Match 0.7%; Score 8; DB 2; Length 300;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 494 ESEILDQ 501
 |||||
 DB 18 ESEILDQ 25

RESULT 10
 093F70 PRELIMINARY; PRT; 300 AA.
 AC 093F70;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE MNXG (FRAGMENT).
 GN MNXG.
 OS Bacillus sp. MB-7.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=161535;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MB-7;
 RA Francis C.A., Tebo B.M.;
 RT "Enzymatic manganese(II) oxidation by metabolically-dormant spore of
 RT diverse Bacillus species.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF326389; AAL30441.1; -.
 FT NON_TER 1
 FT NON_TER 300
 SQ SEQUENCE 300 AA; 33636 MW; DAE6191540BA2761 CRC64;

Query Match 0.7%; Score 8; DB 2; Length 300;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 494 ESEILDQ 501
 |||||
 DB 18 ESEILDQ 25

RESULT 11
 073906 PRELIMINARY; PRT; 474 AA.
 ID 073906;
 AC 073906;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE CDS PROTEIN PRECURSOR.
 GN CDS.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H.B2; TISSUE=THYMUS;
 RX MEDLINE=98250124; PubMed=9590242;
 RA Koskinen R., Goebel T.W.F., Tregaskes C.A., Young J.R., Vainio O.;
 RT "The structure of avian CDS implies a conserved function.";
 RL J. Immunol. 160:4943-4950(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21332904; PubMed=11439160;
 RA Koskinen R., Salomonsen J., Goodchild M., Bumsstead N., Boyd Y.,
 RA Vainio O.;
 RT "Structure and chromosomal localization of chicken CDS.";
 RL Scand. J. Immunol. 54:141-145(2001).
 DR EMBL: Y12011; CAAT7239.1; -.
 DR InterPro: IPR001190; SRCR.
 DR Pfam: PF00530; SRCR; 3.
 DR SMART: SM00202; SR; 3.
 DR PROSITE: PSS0287; SRCR_2; 3.
 KW SIGNAL.
 FT SIGNAL 1
 FT SIGNAL 22
 SQ SEQUENCE 474 AA; 50998 MW; 3AF2B841591AED65 CRC64;

Query Match 0.7%; Score 8; DB 13; Length 474;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 468 GTVASICL 475
 |||||
 DB 358 GTVASICL 365

RESULT 12
 09EV50 PRELIMINARY; PRT; 493 AA.
 ID 09EV50;
 AC 09EV50;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE CATALASE (EC 1.11.1.6).
 GN KATA.
 OS Staphylococcus xylosoy.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=1288;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CURED STRAIN C2A;
 RA Barriere C., Bruckner R., Talon R.;
 RT "Characterization of catalases in Staphylococcus xylosoy.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ295151; CAC14836.1; -.
 DR HSSP: P42321; ZCAE.
 DR InterPro: IPR002226; Catalase.

DR Pfam: PF00199; catalase; 1.
 DR PRINTS: PR00067; CATALASE.
 DR Prodom: PD000510; Catalase; 1.
 DR PROSITE: PS00437; CATALASE_1; 1.
 KW Oxidoreductase; peroxidase.
 SQ SEQUENCE 493 AA; 56874 MW; 900B959518A9404B CRC64;

Query Match 0.7%; Score 8; DB 2; Length 493;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 310 QARNHKDN 317
 |||||
 DB 266 QARNHKDN 273

RESULT 13
 O9KWI9 PRELIMINARY; PRT; 505 AA.

AC 09KWI9; 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE CATALASE.
 OS Staphylococcus warneri.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxId=1292;
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISK-1;
 RA Mizuno K., Fukuda D., Kakiyama M., Kohno M., Ha T.L., Sonomoto K.;
 RT "Purification and Gene Cloning of Catalase from Staphylococcus warneri
 ISK-1."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB045340; BAA97560.1; -
 DR HSPF; P42321; ZCAE.
 DR InterPro: IPR002226; Catalase.
 DR Pfam: PF00199; catalase; 1.
 DR PRINTS: PR00067; CATALASE.
 DR Prodom: PD000510; Catalase; 1.
 DR PROSITE: PS00437; CATALASE_1; UNKNOWN_1.
 DR PROSITE: PS00438; CATALASE_2; 1.
 SQ SEQUENCE 505 AA; 58042 MW; 106AF448457EC1F6 CRC64;

Query Match 0.7%; Score 8; DB 2; Length 505;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 310 QARNHKDN 317
 |||||
 DB 269 QARNHKDN 276

RESULT 14
 O9CD12 PRELIMINARY; PRT; 519 AA.

AC 09CD12; 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)
 DE PUTATIVE SECRETED PROTEIN.
 GN MCEIC OR M2591.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
 OX NCBI_TaxId=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TN;
 RX MEDLINE=21128732; PubMed=11234002;

RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Whingall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 DR EMBL: AL583926; CAC32123.1; -
 DR Leproma: M2591; -
 DR InterPro: IPR003399; Mce.
 DR InterPro: IPR002965; P_rich_extensin.
 DR Pfam: PF02470; mce; 1.
 DR PRINTS: PR01217; PRICHEXTENSIN.
 KW Complete proteome.
 SQ SEQUENCE 519 AA; 54450 MW; 8A3A6AEFAD619EBD CRC64;

Query Match 0.7%; Score 8; DB 16; Length 519;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 561 FPDPMGTR 568
 |||||
 DB 346 FPDPMGTR 353

RESULT 15
 O993G4 PRELIMINARY; PRT; 699 AA.

AC 0993G4; 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE ORF45.
 OS callitriche herpesvirus 3.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxId=106331;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CJ0149;
 RX MEDLINE=21107697; PubMed=11158621;
 RA Cho Y., Ramer J., Rivallier P., Quink C., Garber R.L., Beier D.R.,
 RA Wang F.;
 RT "An Epstein-Barr-related herpesvirus from marmoset lymphomas."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:1224-1229(2001).
 DR EMBL: AF319782; AAK38254.1; -
 SQ SEQUENCE 699 AA; 75289 MW; 89B75E925AE64AB1 CRC64;

Query Match 0.7%; Score 8; DB 12; Length 699;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 409 EYTVANPV 416
 |||||
 DB 398 EYTVANPV 405

Search completed: May 21, 2002, 14:56:10
 Job time: 169 sec

THIS PAGE BLANK (USPTO)